

Autonome Systeme und Arbeit: Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt

Hirsch-Kreinsen, Hartmut (Ed.); Karačić, Anemari (Ed.)

Veröffentlichungsversion / Published Version

Sammelwerk / collection

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
transcript Verlag

Empfohlene Zitierung / Suggested Citation:

Hirsch-Kreinsen, H., & Karačić, A. (Hrsg.). (2019). *Autonome Systeme und Arbeit: Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt*. Bielefeld: transcript Verlag. <https://doi.org/10.14361/9783839443958>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:
<https://creativecommons.org/licenses/by-nc-nd/4.0>

Hartmut Hirsch-Kreinsen,
Anemari Karačić (Hg.)

AUTONOME SYSTEME UND ARBEIT

Perspektiven, Herausforderungen
und Grenzen der Künstlichen Intelligenz
in der Arbeitswelt

[transcript]

Hartmut Hirsch-Kreinsen, Anemari Karačić (Hg.)
Autonome Systeme und Arbeit

HARTMUT HIRSCH-KREINSEN, ANEMARI KARAČIĆ (Hg.)

Autonome Systeme und Arbeit

Perspektiven, Herausforderungen

und Grenzen der Künstlichen Intelligenz in der Arbeitswelt

[transcript]

Dieser Band geht hervor aus der Tagung »Autonome Systeme und Arbeit – Perspektiven, Widersprüche und Konsequenzen«, die am 25. Januar 2018 in Düsseldorf durch das Forschungsinstitut für gesellschaftliche Weiterentwicklung (FGW) e.V. ausgerichtet wurde. Diese Publikation wurde finanziell durch das FGW unterstützt.



Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.



Dieses Werk ist lizenziert unter der Creative Commons Attribution-NonCommercial-NoDerivs 4.0 Lizenz (BY-NC-ND). Diese Lizenz erlaubt die private Nutzung, gestattet aber keine Bearbeitung und keine kommerzielle Nutzung. Weitere Informationen finden Sie unter <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

Um Genehmigungen für Adaptionen, Übersetzungen, Derivate oder Wiederverwendung zu kommerziellen Zwecken einzuholen, wenden Sie sich bitte an rights@transcript-verlag.de

Die Bedingungen der Creative-Commons-Lizenz gelten nur für Originalmaterial. Die Wiederverwendung von Material aus anderen Quellen (gekennzeichnet mit Quellenangabe) wie z.B. Schaubilder, Abbildungen, Fotos und Textauszüge erfordert ggf. weitere Nutzungsgenehmigungen durch den jeweiligen Rechteinhaber.

© 2019 transcript Verlag, Bielefeld

Umschlaggestaltung: Maria Arndt, Bielefeld

Umschlagabbildung: © phonlamaipphoto – Fotolia.com (bearbeitet)

Korrekturat: Tanja Jentsch (7Silben), Bottrop

Satz: Anemari Karačić, Düsseldorf

Druck: Majuskel Medienproduktion GmbH, Wetzlar

Print-ISBN 978-3-8376-4395-4

PDF-ISBN 978-3-8394-4395-8

<https://doi.org/10.14361/9783839443958>

Gedruckt auf alterungsbeständigem Papier mit chlorfrei gebleichtem Zellstoff.

Besuchen Sie uns im Internet: <https://www.transcript-verlag.de>

Bitte fordern Sie unser Gesamtverzeichnis und andere Broschüren an unter: info@transcript-verlag.de

Inhalt

Vorwort | 7

Einleitung

Hartmut Hirsch-Kreinsen und Anemari Karačić | 9

POTENZIALE UND GRENZEN DER ANWENDUNG AUTONOMER SYSTEME

Autonome Systeme

Potenziale und Herausforderungen

Thomas Kuhn und Peter Liggesmeyer | 27

Künstliche Intelligenz in Bildungs- und Arbeitsräumen

Internet der Dinge, Big Data,

Personalisierung und adaptives Lernen

Miloš Kravčák, Carsten Ullrich und Christoph Igel | 47

Grenzen und Widersprüche der Entwicklung und Anwendung ‚Autonomer Systeme‘

Peter Brödner | 69

AUTONOME SYSTEME UND DER WANDEL VON ARBEIT – KONZEPTIONELLE BEITRÄGE UND EMPIRISCHE FORSCHUNGSERGEBNISSE

Autonome Systeme und Künstliche Intelligenz

Herausforderungen für die Arbeitssystemgestaltung

Detlef Gerst | 101

Assimilierende versus komplementäre Adaptivität

Grenzen (teil-)autonomer Systeme

Nobert Huchler | 139

Die Autonomie instrumentell genutzter Technik

Eine handlungstheoretische Analyse

Ingo Schulz-Schaeffer | 181

**Herausforderungen der sozio-technischen Evaluation
der Arbeit mit Autonomen Systemen**

Thomas Herrmann und Jan Nierhoff | 207

**HISTORISCHE UND GESELLSCHAFTLICHE
PERSPEKTIVEN**

Technik und Autonomie

Kulturhistorische Bemerkungen zu einem komplexen Verhältnis

Martina Heßler | 247

Haben autonome Maschinen Verantwortung?

Christoph Hubig | 275

Autorinnen und Autoren | 299

Vorwort

In absehbarer Zukunft werden Autonome Systeme Einzug auch in die Arbeitswelt halten. Ähnlich wie die viel diskutierten selbstfahrenden Autos basieren sie auf den Prinzipien und Methoden der Künstlichen Intelligenz. In der Arbeitswelt handelt es sich um Systeme wie autonome Roboter, adaptive Assistenzsysteme oder auch intelligente Logistiksysteme. Erste Anwendungsfälle von erfolgreich implementierten, zumindest als teilautonom anzusehenden Systemen dieser Art in Industriebetrieben sind verschiedentlich dokumentiert.

Diese Entwicklung stellt die Arbeitsforschung in konzeptioneller und empirischer Hinsicht vor große Herausforderungen. Es werden nicht nur die im Prinzip bekannten Fragen nach den technologischen Konsequenzen für Arbeitsplätze und Qualifikationen aufgeworfen, sondern ungeklärt ist auch, welche neuen Herausforderungen sich damit für Analysekonzepte und Gestaltungskriterien für Arbeit verbinden. Darüber hinaus stellen sich arbeitspolitische, gesellschaftliche und ethische Fragen nach den sozialen Konsequenzen dieser Systeme. Letztlich ist aber auch unklar, welche Entwicklungs- und Anwendungspotenziale Autonome Systeme und Künstliche Intelligenz in der Arbeitswelt tatsächlich haben und wo mögliche Grenzen und Widersprüche ihrer Nutzbarkeit liegen.

Im Rahmen des vorliegenden Bandes wird diesen Fragen aus dem Blickwinkel verschiedener wissenschaftlicher Disziplinen nachfolgend genauer nachgegangen. Die Beiträge basieren in Teilen auf der im Januar 2018 in Düsseldorf durchgeführten Tagung „Autonome Systeme“.

me und Arbeit – Perspektiven, Widersprüche und Konsequenzen“. Organisiert wurde die Tagung vom Forschungsinstitut für Gesellschaftliche Weiterentwicklung e.V. (FGW) im Rahmen des Themenbereichs Digitalisierung von Arbeit/Industrie 4.0. Das FGW hat zum Ziel, auf der Basis wissenschaftlicher Erkenntnisse öffentlich orientierte Dialoge zu gesellschaftspolitisch relevanten Herausforderungen zu initiieren. Entsprechend ist die Zielsetzung des vorliegenden Bandes, einen interdisziplinären Beitrag zum laufenden Diskurs über die wachsende Bedeutung und die Konsequenzen von Systemen der Künstlichen Intelligenz für Arbeit zu erbringen. Jenseits der vielen spekulativen Erwägungen und Diskussionen soll eine fundierte Einschätzung gewonnen werden, welche arbeits- und gesellschaftspolitischen Weichenstellungen für einen verantwortungsvollen Umgang mit den neuen Technologien geboten sind.

Gedankt sei an dieser Stelle allen Autorinnen und Autoren für die pragmatische und fruchtbare Zusammenarbeit bei der Erstellung und Bearbeitung der Beiträge; nicht zuletzt für die schnelle und vorbehaltlose Bereitschaft, das Manuskript innerhalb eines relativ knappen Zeitplans zur Publikationsreife zu bringen. Des Weiteren danken wir Tanja Jentsch für redaktionelle Arbeiten an den Beiträgen und Julia Wicczorek und Johanna Tönsing für die Betreuung des Bandes. Zudem sei auch dem transcript Verlag für sein Interesse an dieser Thematik und der vorliegenden Publikation ausdrücklich gedankt.

Düsseldorf, im Herbst 2018

Die Herausgeber

Einleitung

Hartmut Hirsch-Kreinsen und Anemari Karačić

1. TECHNOLOGIEVERSPRECHEN AUTONOME SYSTEME

Im laufenden Diskurs über die Digitalisierung und ihre Perspektiven für den gesellschaftlichen Wandel wird vielfach dezidiert die Erwartung formuliert, dass digitale Technologien in Zukunft in vielen Anwendungsfeldern als Autonome Systeme ohne menschliche Eingriffe aktionsfähig werden. Es wird davon ausgegangen, dass die schnelle technologische Entwicklung Systeme ermöglicht, die lernfähig sind und in einer im Prinzip unbekannten Situation eigenständig entscheiden sowie agieren können. Solche Systeme weisen grundsätzlich die Fähigkeit auf, komplexe Verarbeitungsketten von Daten, automatische Objektidentifikationen und Sensorfunktionen auf verschiedensten Ebenen bis hin zur Schaffung einer für die jeweilige Zielsetzung des Systems hinreichend genauen digitalen Repräsentation der Wirklichkeit realisieren und beherrschen zu können. Informationstechnologische Voraussetzung hierfür sind die massiven Leistungssteigerungen von Computerhardware und neuere Entwicklungen im Feld der Künstlichen Intelligenz (KI).

Möglich werden damit Systeme, die in der Lage sein sollen, menschenähnlich zu handeln und zu entscheiden und daher auch als ‚intelligent‘ oder ‚smart‘ charakterisiert werden. Als ihr zentrales Merkmal

gilt ihre situationsspezifische Adaption- und Lernfähigkeit, im Englischen als *Machine Learning* oder *Deep Learning* bezeichnet. Informatiker/-innen erwarten, dass damit ein der menschlichen Lernfähigkeit entsprechendes Leistungsniveau der digitalen Maschinen erreicht wird. Insbesondere, so die Perspektive, sollen intelligente Systeme zukünftig in der Lage sein, auch implizites Wissen und menschliches Erfahrungswissen zu erfassen und zu verarbeiten. Denn bislang galten kontextabhängige und auf hohen Anteilen impliziten Wissens basierende Lernprozesse und damit verbundene Aufgaben aufgrund fehlender und nicht explizierbarer Handlungsregeln als grundsätzlich nicht überwindbare Barrieren für Automatisierung bzw. Algorithmisierung. Bezeichnet werden damit Handlungssituationen, die durch Michael Polanyi's (vgl. 1966) bekannte Formel ‚We know more than we can tell‘ charakterisiert werden können. Konkret geht es dabei um Handlungselemente der verschiedensten Art und in den unterschiedlichsten Zusammenhängen, die durch ein hohes Maß an Flexibilität, Urteilsvermögen, sozialer Interaktion und Kommunikation sowie eben auch akkumulierter Erfahrung gekennzeichnet sind. Dieses wissenstheoretisch auch als ‚Polanyi's Paradox‘ bekannte Problem soll nun endlich, so die dezidierte Erwartung von einflussreichen Informatiker/-innen, durch KI-basierte Autonome Systeme überwunden, das heißt automatisiert werden (vgl. insbesondere McAfee/Brynjolfsson 2017: 66ff.).

In der wissenschaftlichen und in der öffentlichen Debatte wird das Thema der Autonomen Systeme inzwischen intensiv und facettenreich diskutiert. Getrieben wird die Diskussion insbesondere durch die Entwicklungen im Automobilbereich und das Thema des autonomen Fahrens, aber auch weitverbreitete Vorstellungen über die Aktionsfähigkeiten menschenähnlicher Roboter. Zudem rückt innovationspolitisch das Thema Autonome Systeme zunehmend in den Fokus. Denn es wird allgemein davon ausgegangen, dass sich mit der Entwicklung und Anwendung solcher Systeme in sehr vielen gesellschaftlichen Bereichen ganz erhebliche Nutzenpotenziale verbinden und damit das deutsche Innovationssystem im internationalen Wettbewerb deutlich gestärkt werden könnte. Daher sollten erhebliche Innovationsanstrengungen

eingeleitet werden, um insbesondere die schnelle Weiterentwicklung der KI als technologische Basis für Autonome Systeme zu fördern (vgl. Bundesministerium für Wirtschaft und Energie 2018; EFI – Expertenkommission Forschung und Innovation 2018). Die Künstliche Intelligenz wird in diesem Rahmen oftmals als ‚Schlüsseltechnologie‘ betrachtet, in die zahlreiche Hoffnungen und Erwartungen projiziert werden.

Im Hinblick auf die Möglichkeiten und Anwendungspotenziale Autonomer Systeme werden im Kontext der laufenden Digitalisierungsdebatte geradezu spektakuläre ökonomische Gewinne und die Lösung vielfältiger gesellschaftlicher Herausforderungen prognostiziert und postuliert. Den Schätzungen einer EU-Studie zufolge kann insgesamt mit Marktpotenzialen Autonomer Systeme von mehreren 100 Milliarden US-Dollar in den nächsten Jahren ausgegangen werden (vgl. Platforms4CPS 2017). Neben dem autonomen Fahren werden als zentrale Anwendungsbereiche die Gestaltung intelligenter Gebäude, sogenannte *Smart Homes*, die Bewältigung von Anforderungen menschenfeindlicher Umgebungen, wie sie beispielsweise beim Rückbau von Kernkraftanlagen auftreten, und insbesondere auch die industrielle Produktion und die dort anzutreffenden Arbeitsprozesse hervorgehoben (vgl. Fachforum Autonome Systeme im Hightech-Forum 2017).

In der Industrie können Autonome Systeme beispielsweise im Rahmen von fahrerlosen Transportsystemen für die unternehmensinterne Logistik oder bei Mensch-Roboter-Kollaborationen zum Einsatz kommen. Sie sollen damit all jene aus dem Industrie-4.0-Diskurs bekannten Zielsetzungen realisieren: eine gleichzeitige Beschleunigung und Flexibilisierung der Produktion sowie ein höherer Grad an Individualisierung der Produkte. Zudem sollen Autonome Systeme Ausfallzeiten durch vorausschauende Wartung, sogenannte Predictive Maintenance reduzieren (vgl. ebd.). Schließlich wird die Steigerung der Ressourceneffizienz der industriellen Produktion durch sich selbstoptimierende Systeme erwartet, um die wachsende Ressourcenknappheit zu bewältigen. Im Hinblick auf die Entwicklung von Arbeit wird unterstrichen, dass die neuen Systeme geradezu zwangsläufig zu sicheren

Arbeitsplätzen, guter Arbeit, anspruchsvollen Jobs und einer deutlichen Verbesserung der Work-Life-Balance führen werden. Auf der Basis Autonomer Systeme, etwa dem Einsatz smarter und adaptiver Robotersysteme eröffneten sich danach zugleich Möglichkeiten, den Fachkräftemangel bewältigen und längerfristig auch die demografischen Probleme beherrschbar machen zu können. Als eine zentrale Voraussetzung hierfür werden die Möglichkeiten der intelligenten Technologien gesehen, Bildungs- und Ausbildungsprozesse zu effektivieren und vor allem durch ihre Adaption an unterschiedliche Qualifikationsniveaus diese auch zielgruppenspezifisch ausgestalten zu können (vgl. Barner et al. 2015).

Diese Erwartungen und Prognosen legen es nahe, digitale Technologien und insbesondere KI-basierte Autonome Systeme innovations-theoretisch als *Promising Technology* zu verstehen. Mit diesem Konzept wird nach den Verlaufsmustern von Technikentwicklung und nach den sozio-ökonomischen Konstitutionsbedingungen neuer Technologiefelder gefragt (vgl. Bender 2005; van Lente/Rip 1998). Anders formuliert: Auch beim Konzept der Autonomen Systeme handelt es sich um ein Technologieversprechen, das vor allem ausgeprägt technikoptimistische Züge im Hinblick auf seine sozio-ökonomischen Konsequenzen aufweist. Damit steht die laufende Digitalisierungsdebatte in der Tradition früherer technikzentrierter Diskurse, wie etwa die über die Kernenergie, die Weltraumfahrt oder die Informationsgesellschaft, die ebenfalls weitreichende positive gesellschaftliche Konsequenzen in Aussicht stellten.

2. WIDERSPRÜCHE UND BARRIEREN

Inwieweit sich diese mit Autonomen Systemen verbundenen Erwartungen und Versprechen problemlos durchsetzen, ist gleichwohl eine offene Frage. Denn aus sozialwissenschaftlicher Sicht lässt sich einwenden, dass diese technikzentrierte Perspektive mit ihrem teilweise weitreichenden Prognoseanspruch zu kurz greift. Denn sowohl die so-

zialwissenschaftliche Innovationsforschung als auch die soziologische Arbeitsforschung verfügen über einen breiten Fundus konzeptioneller und empirischer Forschungsergebnisse, die aufzeigen, dass die Entwicklung und die Diffusion neuer Technologien alles andere als bruchlos und widerspruchsfrei verlaufen und daher die sozio-ökonomischen Effekte kaum eindeutig aus den Potenzialen neuer Technologien ableitbar sind. So betont seit langem die Innovationsforschung, dass Innovationen zwar stets zielgerichtet und dynamisch verlaufen, ihr Verlauf zugleich jedoch risikoreich ist und ihre Effekte ungewiss sind. Entscheidend für die sich jeweils einspielenden Verlaufsmuster von Innovationen und die dadurch angestoßenen strukturellen Veränderungen sind danach zum einen die Erarbeitung und Variation technologisch möglicher Entwicklungspotenziale, zum anderen ihre Selektion im Lichte von Anforderungserfordernissen und Vermarktungschancen neuer Technologien.

Mit anderen Worten: Übersehen werden bei den weitreichenden technologisch begründeten Erwartungen die in der Regel die Verbreitung der neuen Technologien beeinflussenden realen ökonomischen und sozialen Bedingungen. Dieser Zusammenhang lässt sich instruktiv verdeutlichen, wenn man Joseph Schumpeters begriffliche Differenzierung von Innovationen aufgreift. Danach umfasst ein Innovationsprozess vier Stufen: Invention, Innovation, Diffusion und Implementation bzw. Imitation. Der aktuelle Digitalisierungsdiskurs fokussiert sich zumeist auf die Phasen Invention und Innovation, das heißt der Entwicklung einer Erfindung zu einem marktgängigen Produkt. Entscheidend für die Form der Nutzung neuer Technologien und den Wandel von Arbeit ist jedoch die Frage, wie der *Diffusions- und Implementationsprozess* neuer Technologien und ihre Anpassung an die je gegebenen sozialen und ökonomischen Realitäten verläuft und welche Konsequenzen sich aus diesem Prozess ergeben (vgl. zusammenfassend Rogers 2003). Es sind daher nicht ausschließlich die technologischen Potenziale und Eigenschaften einer Innovation, die bestimmte soziale und ökonomische Auswirkungen erzeugen. Vielmehr gilt:

„[...] [T]he really interesting aspect of new technologies is whether they prompt investors, companies, labor, and markets to change, or whether these factors and organizations of production resist the absorption of new technologies.“ (Erixon/Weigel 2016: 13)

Auf diese Zusammenhänge verweisen im laufenden Diskurs über Digitalisierung und insbesondere die Perspektiven der KI eine ganze Reihe kritischer und skeptischer Argumente. Im Hinblick auf die Perspektiven Autonomer Systeme in industriellen Arbeitsprozessen seien die folgenden hervorgehoben:

Erstens sind nicht alle als autonom bezeichneten Systeme tatsächlich im oben definierten Sinn intelligent und autonom. Vielmehr verweisen sowohl die einschlägige Forschung als auch die Anwendungspraxis auf sehr verschiedene Stufen der Autonomisierung von Produktionstechnik. So weisen die heutigen in den Unternehmen anzutreffenden digitalen Systeme, etwa Assistenzsysteme und vernetzte Anlagen, nur sehr begrenzte Autonomiegrade auf. Weitergehende Entwicklungsschritte in Richtung technologischer Fähigkeiten der Selbstorganisation und Selbstoptimierung seien sehr aufwändig und es wird betont, dass dafür vielfältige komplexe Herausforderungen bewältigt werden müssen (vgl. Damm/Kalmar 2017; EFI – Expertenkommission Forschung und Innovation 2018). So heben Unternehmensvertreter/-innen vielfach hervor, dass entgegen aller Erwartungen von Entwickler/-innen, beispielsweise smarte Roboter in Montageprozessen absehbar nur begrenzt in der Lage sein werden, die zunehmende Teilevarianz, spezifische hochflexible Materialien und den sogenannten ‚Griff in die Kiste‘ zu bewältigen.

Zweitens werden die prognostizierten weitreichenden ökonomischen Effekte vor allem von Betriebspraktiker/-innen deutlich angezweifelt und es wird vielfach gefragt, welchen Nutzen neue digitale Systeme wirklich bieten. So wird auf die – in der laufenden Diskussion nicht systematisch thematisierten – hohen Investitions- und Implementationskosten der neuen Systeme und letztlich unklare Rentabilitätsaussichten hingewiesen (vgl. agiplan/Fraunhofer IML/Zenit 2015). Es ist

davon auszugehen, dass viele Unternehmen, insbesondere kleinere und mittlere Unternehmen, ein verschiedentlich negativ bewertetes Verhältnis zwischen dem hohen prognostizierten Investitionsbedarf und dem daraus resultierenden Umsatzwachstum zögern lässt. Schließlich wird in vielen Betrieben immer wieder auf die unklaren Folgeinvestitionen hingewiesen, die notwendig seien, um die Systeme wirklich zum Laufen zu bringen.

Drittens wird vielfach die Beherrschbarkeit neuer Autonomer Systeme angezweifelt. So zeigen die Arbeitswissenschaften und die arbeitspsychologische Forschung, dass mit fortschreitender Automatisierung und der damit verbundenen Steigerung der Komplexität der Systeme oftmals nur mehr eine begrenzte Beherrschbarkeit der Technologien, damit ein hohes funktionales und ökonomisches Störpotenzial und unkalkulierbare Anforderungen an das Arbeitshandeln einhergehen (vgl. zusammenfassend Grote 2018). Den Forschungsergebnissen zufolge sind effektive Arbeitseingriffe in Systemabläufe, um Störungen präventiv zu vermeiden oder zu beheben, nicht immer möglich, da Beschäftigte oftmals nicht (mehr) in der Lage sind, Autonome Systeme wirksam zu kontrollieren und damit die Verantwortung über den Systembetrieb zu übernehmen. Als zentrale Ursache hierfür gilt, dass aufgrund einer informationstechnologisch begründeten hohen funktionalen und informationellen Distanz zum Systemablauf das Personal kein hinreichend valides Wissen und keine ausreichenden Erfahrungen über relevante Systemfunktionen akkumulieren kann. Gerade auch unter den Bedingungen Autonomer Systeme gelten daher die seit langem diagnostizierten ‚Ironien der Automatisierung‘ (vgl. Bainbridge 1983): Demnach erzeugen hochautomatisierte Prozesse aufgrund ihres ausgeprägten Routineablaufs bei unerwartet auftretenden (in der Regel aber unvermeidbaren) Störungen nur schwer zu bewältigende Arbeitssituationen. Denn zur Störungsbewältigung seien Qualifikationen und Kompetenzen, insbesondere Erfahrungswissen erforderlich, die oftmals im informatisierten Routinebetrieb weder aufgebaut noch auf Dauer erhalten werden könnten. Hinzu kommt, dass die Entscheidungsfindung Autonomer Systeme nicht transparent und demzufolge nicht

nachvollziehbar vonstattengeht, was für den Arbeitsprozess zur Folge haben kann, dass Eingriffe in das System nur begrenzt möglich sind.

Im Hinblick auf die sozialen Effekte der neuen Systeme finden sich darüber hinaus oft negative, ja geradezu dystopische Prognosen, wie sie seit Längerem generell im Digitalisierungsdiskurs formuliert werden. Verwiesen wird auf mögliche hohe Arbeitsplatzverluste in manuellen und kognitiven Arbeitsbereichen, Gefahren der Dequalifizierung sowie ein deutlich erhöhtes Kontrollpotenzial, einer forcierten Flexibilisierung und Prekarisierung sowie wachsenden Belastungen bei der Arbeit (vgl. zusammenfassend Hirsch-Kreinsen/Ittermann/Niehaus 2018). Dabei wird vor allem auch die wachsende Autonomie der digitalen Systeme vielfach kritisch diskutiert. Die Befürchtung ist, dass intelligente Robotersysteme oder selbstfahrende Automobile in Bereiche eindringen, die bislang allein unter der Kontrolle menschlichen Handelns und menschlicher Verantwortlichkeiten standen und dabei völlig außer Kontrolle geraten. Gesehen wird die Gefahr,

„[...] that the rich traditions of moral thought that guide human relationships have no equivalent when it comes to robot-human-interaction [...]. We face a future in which robots will test the boundaries of our ethical and legal frameworks with increasing audacity.“ (Nourbakhsh 2015: 23f.)

Angesprochen werden damit verbundene rechtliche Probleme, insbesondere unklare Haftungsansprüche. Darüber hinaus aber geht es auch um bislang kaum überschaubare, geschweige denn gelöste ethische Probleme, wie die Frage nach einer moralischen Fundierung und Rechtfertigung der von Maschinen getroffenen Entscheidungen oder die Frage, wer letztlich die Verantwortung für ‚Maschinenentscheidungen‘ trägt.

3. DIE BEITRÄGE DES BANDES

Insgesamt sind jenseits aller technikoptimistischen Erwartungen und skeptischen Positionen Richtung, Intensität und Reichweite des durch Autonome Systeme induzierten sozialen und ökonomischen Wandels bislang nur schwer absehbar. Vor allem ist hervorzuheben, dass – wie im Digitalisierungsdiskurs immer wieder betont – die Formen der Nutzung der neuen Technologien und der damit einhergehende sozio-ökonomische Wandel als politisches und unternehmerisches Gestaltungsprojekt zu verstehen sind. Dies gilt insbesondere für industrielle Arbeitsprozesse, auf die sich besonders der Digitalisierungsdiskurs in Deutschland bezieht (vgl. Hirsch-Kreinsen/Ittermann/Niehaus 2018).

Die Beiträge des vorliegenden Sammelbandes zielen daher darauf, auf der Basis des derzeitigen Wissensstandes sowohl die Entwicklungsoptionen Autonomer Systeme als auch die damit verbundenen Konsequenzen für Arbeit auszuloten und Perspektiven zu skizzieren. Im Einzelnen umfasst der Band informatikorientierte, sozialwissenschaftliche, technikhistorische und technikphilosophische Beiträge. Entsprechend ist der Band in drei größere Abschnitte gegliedert.

Im ersten Abschnitt werden die *Potenziale und Grenzen der Anwendung Autonomer Systeme* aus informatikorientierter Perspektive beleuchtet. Der erste Beitrag von *Peter Liggesmeyer* und *Thomas Kuhn* „Autonome Systeme – Potenziale und Herausforderungen“ gibt zunächst einen Überblick über die Einsatzmöglichkeiten sowie Anwendungsgebiete Autonomer Systeme. Neben wirtschaftlichen Potenzialen werden auch mögliche gesellschaftliche Potenziale des Einsatzes Autonomer Systeme beleuchtet. Um jedoch die Potenziale Autonomer Systeme besser ausschöpfen zu können, formulieren die Autoren unterschiedliche Voraussetzungen und Herausforderungen, die im Hinblick auf Interoperabilität, Normierungen von Standards, Definition von Schnittstellen, Integration von Diensten, Sicherheitsstandards etc. noch zu bewältigen sind. Hervorgehoben wird bei der Bewältigung der noch anstehenden Herausforderungen insbesondere die zentrale Rolle von digitalen Zwillingen als ‚Wegbereitern für Autonome Systeme‘, die

Simulationen im virtuellen Raum unter realen Bedingungen ermöglichen und damit die Konsequenzen unterschiedlicher Ausführungsvarianten vorhersehbar gestalten sollen.

Eine konkrete Perspektive auf die Anwendungsmöglichkeiten Autonomer Systeme verfolgt der darauffolgende Beitrag von *Miloš Kravčik*, *Carsten Ullrich* und *Christoph Igel* zum Thema „Künstliche Intelligenz in Bildungs- und Arbeitsräumen: Internet der Dinge, Big Data, Personalisierung und adaptives Lernen“. Die Autoren fokussieren dabei die Möglichkeiten der Nutzung Künstlicher Intelligenz bei der Optimierung von Lernprozessen (*Smart Learning*). Sie erörtern, wie diese eingesetzt werden können, um Lernprozesse zu personalisieren, indem sich die Systeme beispielsweise an die Nutzer/-innen und ihre Bedarfe anpassen können. Am Beispiel des APPsist-Systems wird eine intelligente, adaptive Lerntechnologie vorgestellt, welche sensorgestützt angepasste Hilfestellungen im Hinblick auf Wissensvermittlung für Beschäftigte bereitstellt.¹ Die neuen Technologien sollen vor allem auch dafür genutzt werden, neue Lernformen zu ermöglichen und informelles Lernen am Arbeitsplatz zu fördern.

Demgegenüber rückt der letzte Beitrag des Abschnitts von *Peter Brödner* die „Grenzen und Widersprüche der Entwicklung und Anwendung Autonomer Systeme“ ins Zentrum der Argumentation. Anhand konkreter Beispiele werden die Funktionsweisen (z.B. *Deep Learning*, künstliche neuronale Netze) intelligenter Systeme veranschaulicht und hinsichtlich der damit verbundenen Erwartungen überprüft. Im Hinblick auf den Entwicklungsprozess wird herausgearbeitet, dass nicht die Intelligenz der Systeme für deren Leistungsfähigkeit verantwortlich ist, sondern jene der Entwickler/-innen. Bezogen auf die Anwenderseite wird an der Stelle insbesondere die Problematik der Intransparenz hervorgehoben, da für die Nutzer/-innen die Prozesse der Autonomeren Systeme undurchschaubar und damit unvorhersehbar sind. Plädiert wird abschließend für einen reflektierten Umgang mit den Be-

1 Vgl. <https://edtec.dfki.de/projekt/appsist/> vom 28.9.2018.

grifflichkeiten und Zuschreibungen, die für sogenannte Autonome Systeme zur Verwendung kommen.

Der folgende Abschnitt thematisiert aus sozialwissenschaftlicher Perspektive den Zusammenhang zwischen *Autonomen Systemen und dem Wandel der Arbeit*. Einen allgemeinen sozial- und arbeitswissenschaftlichen Überblick bietet der erste Beitrag von *Detlef Gerst* mit dem Titel „Autonome Systeme und Künstliche Intelligenz: Herausforderungen für die Arbeitssystemgestaltung“. Erörtert wird insbesondere, welche möglichen Konsequenzen der Einsatz dieser Systeme für die Arbeitswelt mit sich bringt und wirft dabei Fragen nach Kontrolle, (De-)Qualifizierung und Gestaltungsmacht auf, die durch den Einsatz intelligenter Technologien neu verhandelt werden müssen. Im Rückgriff auf das etablierte Konzept des sozio-technischen Systems wird argumentiert, dass technische Entwicklungen nur ein Element der Arbeitsgestaltung sind, welches die anderen Elemente in ihrer Gestaltbarkeit nicht vernachlässigen könne. Reflektiert werden im Anschluss an das Thema der Gestaltbarkeit auch arbeitspolitische Fragen, die sich durch den Einsatz Autonomer Systeme und intelligenter Technik stellen.

Auch der arbeitssoziologische Beitrag von *Norbert Huchler* zum Thema „Assimilierende versus komplementäre Adaptivität. Grenzen (teil-)autonomer Systeme“ knüpft an das Konzept der sozio-technischen Systeme an und fokussiert dabei die Frage der Gestaltung und Einbettung (teil-)autonomer Systeme in den Arbeitsprozess. Ausgehend von der Prämisse, dass das zentrale Kriterium der 4.0-Debatte das Bestreben einer zunehmend dezentralen Steuerung und Adaptivität als Voraussetzung für Flexibilität ist, wird theoriegeleitet erörtert, wie (teil-)autonome Systeme gestaltet werden können. Aufgezeigt werden hierbei die Grenzen der Adaptivität und damit einhergehende Konsequenzen im Falle der Nicht-Beachtung dieser Grenzen. Deutlich wird, dass anstelle der gegenwärtigen Tendenz zu einer assimilierenden Adaptivität, welche den Nachteil mit sich bringt, die Stärken menschlichen (Arbeits-)Handelns und damit auch die Grenzen der digitalen Adaptivität zu übergehen, eine komplementäre Adaptivität zu entwi-

ckeln wäre, welche eine situative und flexible Koordination der Handlungsträgerschaft zwischen Mensch und technischem System ermöglicht. Letztlich ist der Beitrag auch als ein Plädoyer zu verstehen, die Position des ‚Menschen im Mittelpunkt‘ in der 4.0-Debatte nicht nur normativ zu postulieren, sondern auch funktional zu berücksichtigen.

Der nachfolgende konzeptionell orientierte techniksoziologische Beitrag von *Ingo Schulz-Schaeffer* mit dem Titel „Die Autonomie instrumentell genutzter Technik. Eine handlungstheoretische Analyse“ untersucht am Gegenstand der Nutzung von Technik die Frage nach der Verteilung von menschlicher Kontrolle und technischer Autonomie in unterschiedlichen Handlungsdimensionen. Der Autor zeichnet nach, wie die Delegation von Handlungen an Technik sich auf Handlungsentwurf und Handlungsziele auswirken kann und nimmt hierfür drei Formen instrumenteller Nutzung von Technik in den Blick, die unterschiedlichen Autonomiegraden der Technik entsprechen. Dabei zeigt sich, dass die Steuerung bzw. Kontrolle einer Handlung durchaus von den dafür eingesetzten technischen Mitteln beeinflusst wird und im Falle des Einsatzes von technischen Mitteln mit besonderen Kompetenzen, als Beispiel für einen größeren Anteil an Autonomie der Technik, auch die Frage nach dem Einfluss der eingesetzten Mittel auf die Ziele der Handlung an Relevanz gewinnt.

Der letzte Beitrag des zweiten Teils des Bandes befasst sich aus arbeitswissenschaftlicher Perspektive mit den „Herausforderungen der sozio-technischen Evaluation der Arbeit mit Autonomen Systemen“. *Thomas Herrmann* und *Jan Nierhoff* diskutieren die Frage, wie Heuristiken genutzt werden können, um den Einsatz Autonomer Systeme im Rahmen einer sozio-technischen Systemgestaltung zu optimieren. Am Gegenstand eines eigens entwickelten Heuristik-Sets wird veranschaulicht, wie dieses für den Aufbau eines ganzheitlichen Systemverständnisses genutzt werden kann. Die Autoren verdeutlichen hierbei, dass die Rolle des Menschen sich nicht zwingend durch eine höhere Autonomie des technischen Systems hin zu weniger menschlicher Autonomie bewegen muss, sondern menschliche Autonomie durchaus im Hinblick auf koordinative Aufgaben relevant bleibe. Menschliche In-

tervention könne demnach auch verstanden werden als ein Beitrag zur Weiterentwicklung und Rekonfiguration Autonomer Systeme.

Schließlich widmet sich der dritte Abschnitt den *historischen und gesellschaftlichen Perspektiven* im Umgang mit Autonomen Systemen. Ziel dieses Abschnitts ist eine Reflexion der gesellschaftlichen Implikationen der Einführung Autonomer Systeme. In ihrem technikhistorischen Beitrag „Technik und Autonomie. Kulturhistorische Bemerkungen zu einem komplexen Verhältnis“ rückt *Martina Heßler* den historischen Wandel der Wahrnehmung des Verhältnisses von Autonomie und Technik ins Zentrum. Am Beispiel unterschiedlicher Debatten im 20. und 21. Jahrhundert geht die Autorin insbesondere zwei Fragen nach: Wie wandelte sich der Autonomiebegriff im Verhältnis zur Technik? Und welche Position nahm und nimmt der Mensch darin jeweils ein? Entlang dieser Analysen wird deutlich, wie unterschiedlich Autonomie jeweils verstanden wurde und wird und mit welchen verschiedenen Auffassungen für die Position des Menschen im Verhältnis zur Technik dies jeweils verbunden ist.

Abschließend erörtert der Technikphilosoph *Christoph Hubig* die Frage „Haben autonome Maschinen Verantwortung?“. Autonomie suggeriere zunächst ein eigenverantwortliches Handeln, welches in der Regel mit der Verantwortungsübernahme für ein solches in Verbindung gebracht wird. Doch wie verhält es sich mit Autonomen Systemen? Was bedeutet Verantwortlichkeit in diesem Fall? Diesen Fragen geht der Beitrag nach, indem zunächst Verantwortlichkeit in ihren unterschiedlichen Dimensionen und Facetten reflektiert wird. Daraufhin werden Kriterien für Verantwortlichkeit entwickelt, die im weiteren Verlauf des Beitrags eingehender analysiert werden, und Leitdifferenzen zwischen Mensch und Maschine herausgearbeitet.

Insgesamt zeigt sich in den Beiträgen zu den Besonderheiten und sozialen Konsequenzen Autonomer Systeme, dass ein Verständnis von technischen Systemen als bloßes Handlungsobjekt zu kurz greift und daher einmal mehr die Frage nach dem absehbaren Muster verteilter Handlungsträgerschaft zwischen Maschine und Mensch relevant wird. Dabei bleibt jedoch bis heute die zentrale Frage nach der tatsächlichen

und im Arbeitsprozess nutzbaren Autonomie der Maschinen offen. Ersichtlich wird in den Beiträgen zudem, dass es insbesondere einer breiten gesellschaftlichen Debatte bedarf, wie die neue Arbeitsteilung zwischen Menschen und technischen Systemen definiert wird, das heißt welche Aufgaben und Tätigkeiten letztlich wirklich Autonomen Systemen überlassen werden sollen bzw. können und welche weiterhin von Menschen ausgeführt werden; sei es aus Gründen der Verantwortlichkeit oder bedingt durch Fragen der Kompetenz. Deutlich wird aber auch, dass derzeit weder technikoptimistische noch ausgesprochen pessimistische Erwartungen im Hinblick auf die arbeits- und gesellschaftspolitischen Konsequenzen dieser Systeme hinreichend begründbar sind. Vielmehr sind wohl auf absehbare Zeit die mit dieser Technologie verbundene Richtung, Intensität und Reichweite des sozialen und ökonomischen Wandels nur schwer absehbar.

LITERATUR

- agiplan/Fraunhofer IML/Zenit (2015): Erschließen der Potenziale der Anwendung von ‚Industrie 4.0‘ im Mittelstand, Studie im Auftrag des BMWI, Mülheim a.d.R.: agiplan.
- Bainbridge, Lisanne (1983): „Ironies of automation“, in: *Automatica* 19/6, S. 775-779.
- Barner, Andreas/Neugebauer, Reimund/Stratmann, Martin/Veit, Eberhard (2015): Innovationspotenziale der Mensch-Maschine-Interaktion, Dossier, Berlin: acatech.
- Bender, Gerd (2005): „Technologieentwicklung als Institutionalisierungsprozess“, in: *Zeitschrift für Soziologie* 34/3, S. 170-187.
- Bundesministerium für Wirtschaft und Energie (2018): Eckpunkte der Bundesregierung für eine Strategie Künstliche Intelligenz, https://www.bmwi.de/Redaktion/DE/Downloads/E/eckpunktepapier-ki.pdf?__blob=publicationFile&v=4 vom 18.7.2018.
- Damm, Werner/Kalmar, Ralf (2017): „Autonome Systeme“, in: *Informatik-Spektrum* 40/5, S. 400-408.

- EFI – Expertenkommission Forschung und Innovation (2018): Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands 2018, Berlin: EFI– Expertenkommission Forschung und Innovation.
- Erixon, Fredrik/Weigel, Björn (2016): *The Innovation Illusion*, New Haven/London: Yale University Press.
- Fachforum Autonome Systeme im Hightech-Forum (2017): *Autonome Systeme – Chancen und Risiken für Wirtschaft, Wissenschaft und Gesellschaft*, Langfassung des Abschlussberichts, Berlin, http://www.hightech-forum.de/fileadmin/PDF/autonome_systeme_abschlussbericht_langversion.pdf vom 18.9.2018.
- Grote, Gudela (2018): „Gestaltungsansätze für das komplementäre Zusammenwirken von Mensch und Technik in Industrie 4.0“, in: Hartmut Hirsch-Kreinsen/Peter Ittermann/Jonathan Niehaus (Hg.), *Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen*, 2. aktualisierte und überarbeitete Aufl., Baden-Baden: Nomos, S. 215-232.
- Hirsch-Kreinsen, Hartmut/Ittermann, Peter/Niehaus, Jonathan (Hg.) (2018): *Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen*, 2. aktualisierte und überarbeitete Aufl., Baden-Baden: Nomos.
- McAfee, Andrew/Brynjolfsson, Erik (2017): *Machine – Platform – Crowd: Harnessing our Digital Future*, New York/London: W. W. Norton & Company.
- Nourbakhsh, Illah Reza (2015): „The Coming Robot Dystopia“, in: *Foreign Affairs* 94/4, S. 23–28.
- Platforms4CPS (2017): *Market Segmentation for CPS Technology. Deliverable 1.1, Version 01.0*, https://www.platforms4cps.eu/fileadmin/user_upload/D1.1_Market_segmentation_for_CPS_technology.pdf vom 18.7.2018.
- Polanyi, Michael (1966): *The Tacit Dimension*, New York: Anchor Books.
- Rogers, Everett M. (2003): *Diffusion of Innovations*, 5. Aufl., New York: Free Press.

van Lente, Harro/Rip, Arie (1998): „Expectations in Technological Developments: An Example of Prospective Structures to be Filled in by Agency“, in: Cornelis Disco/Barend van der Meulen (Hg.), *Getting New Technologies Together. Studies in Making Socio-technical Order*, Berlin: de Gruyter, S. 203-229.

Potenziale und Grenzen der Anwendung Autonomer Systeme

Autonome Systeme

Potenziale und Herausforderungen

Thomas Kuhn und Peter Liggesmeyer

AUTONOME SYSTEME – POTENZIALE UND GRENZEN

Ein System ist dann autonom, wenn es ohne menschliches Zutun und ohne eine detaillierte Programmierung für eine Situation ein vorgegebenes Ziel selbständig und an seine Situation angepasst erreichen kann. Autonome Systeme unterscheiden sich daher von heute existierenden eingebetteten Systemen durch ihre Fähigkeit, eigenständig Entscheidungen zu treffen – auch dann, wenn diese nicht detailliert programmiert wurden. Jedes Autonome System verfolgt ein oder mehrere Ziele. Diese sind in der Programmierung des Systems festgelegt. Das autonome Fahrzeug wird z.B. seine Passagier/-innen mittels einer optimalen Route zu ihrem Ziel bringen. Die Definition der optimalen Route hängt dabei von den Präferenzen der Passagier/-innen ab: es kann die schnellste Route sein, die Route, die am wenigsten Kraftstoff verbraucht, oder auch die Route mit der schönsten Aussicht. Das autonome Fahrzeug übernimmt dabei sowohl die unmittelbaren Fahrfunktionen, das heißt das Lenken, Beschleunigen und Bremsen, als auch strategische Funktionen wie die Wahl der Fahrspur, die Entscheidung für eine Route und gegebenenfalls die Auswahl einer alternativen Route

im Fall einer Verkehrsstörung. Das Potenzial des Autonomen Systems liegt in diesem Fall zunächst in der Entlastung der Passagier/-innen, die, anstatt das Fahrzeug zu fahren, anderen Tätigkeiten nachgehen können.

Die potenziellen Anwendungsgebiete von Autonomen Systemen sind jedoch viel größer. In allen Bereichen, in denen Menschen nicht oder nicht wirtschaftlich arbeiten können oder wollen, können und werden Autonome Systeme zukünftig anfallende Arbeiten übernehmen. In der Produktionstechnik wird eine autonome Fertigungsplanung selbständig entscheiden, wann und unter Nutzung welcher Maschinen ein bestimmtes Produkt am besten gefertigt werden kann. Autonom fahrende Lastwagen benötigen keine Fahrer/-innen mehr, sondern folgen einem Führungsfahrzeug auf der Autobahn. Bereits heute gibt es autonome Schienenfahrzeuge, die ohne Fahrer/-innen funktionieren und dennoch ihre Passagier/-innen sicher zum Ziel bringen. Autonome Fahrzeuge werden eigene Ziele gegenüber übergeordneten Systemzielen abwägen. Ähnlich wie menschliche Fahrer/-innen wird ein autonomes Fahrzeug entscheiden können, ob es in einer Stausituation einem Fahrzeug einer nicht vorfahrtsberechtigten Straße Vorfahrt gewährt und so das Systemziel ‚Fortkommen für alle Fahrzeuge‘ dem eigenen Ziel der möglichst schnellen Zielerreichung überordnet.

Trotz dieser Fähigkeit des Abwägens und des Treffens von Entscheidungen sind Autonome Systeme immer an einen Kontext gebunden. Sie treffen innerhalb dieses Kontexts Entscheidungen, die durchaus komplex sein können. Diese Entscheidungen treffen sie anhand definierter Regeln, die auf Variablen und numerischen Werten beruhen. Die Entscheidungsfindung basiert auf definierten Algorithmen, unabhängig davon, ob diese als künstliche neuronale Netze, als fest programmierte Regeln oder als Optimierungsmodelle vorliegen. Autonome Systeme sind daher überall dort einsatzfähig, wo Entscheidungen anhand von messbaren Kriterien getroffen werden. Autonome Systeme sind jedoch nicht kreativ und können keine Entscheidungen treffen, für die sie nicht vorbereitet wurden. Auch in Zukunft werden Autonome

Systeme den Rahmen ihrer Programmierung daher nicht verlassen können.

Das Fehlen dieser Fähigkeit definiert eine Grenze der Einsatzpotenziale Autonomer Systeme. Kreative Entscheidungen, also das sinnvolle Reagieren auf vollkommen neue Situationen, werden auch in Zukunft dem Menschen vorbehalten bleiben. Auf gänzlich unbekannte Situationen wird auch ein Autonomes System nicht reagieren können. Beispielhaft kann man dies anhand der folgenden Situation diskutieren. Im Normalfall wird ein autonomes Fahrzeug eine Autobahn auf einer Fahrspur nur in einer definierten Richtung befahren. Kommt es zu einem Stau, wird ein autonomes Fahrzeug anhalten, und erst dann weiterfahren, wenn vor ihm genügend Platz entstanden ist. Ist eine Fahrspur z.B. aufgrund von Räumarbeiten gesperrt, so kann das autonome Fahrzeug auf eine andere Fahrspur ausweichen, da ihm das Konzept des Spurwechsels bekannt ist. Selten kommt es jedoch vor, dass die Polizei entscheidet, die wartenden Fahrzeuge z.B. aufgrund einer vorhersehbar langen Sperrung umdrehen zu lassen, und die Fahrbahn entgegen der Fahrtrichtung bis zur nächsten Ausfahrtmöglichkeit zu befahren. Wurde diese Möglichkeit nicht in das autonome Fahrzeug programmiert, wird es diese Entscheidung von sich aus nicht treffen können und auch die entsprechende Anweisung der Polizei nicht verstehen. Hier stößt das Autonome System daher an die Grenzen seiner Programmierung, die es selbständig nicht überwinden kann.

Wirtschaftliche Potenziale Autonomer Systeme

Das wirtschaftliche Potenzial von Autonomen Systemen liegt in deren Fähigkeit, selbständig und effizient Entscheidungen treffen zu können. Menschen müssen diese Systeme nicht mehr zu allen Zeiten leiten. Dies führt zu effizienteren Abläufen und zu einer verbesserten Ressourcennutzung.

Ein naheliegendes Beispiel ist die Produktion. Viele Produktionslinien sind heute auf die Massenproduktion von gleichen Produkten optimiert. Dabei ist der Gewinn pro gefertigtem Produkt häufig sehr ge-

ring; um wirtschaftlich zu arbeiten, müssen daher hohe Stückzahlen der gleichen Produkte produziert werden. Die Produktion von Spezialanfertigungen würde prozentual deutlich höhere Gewinne pro Stück ermöglichen. Die Umsetzung einer solchen Spezialanfertigung scheitert jedoch häufig an den nur geringen Stückzahlen und an dem absolut geringen Umsatz, der in keinem Verhältnis zu den Rüstkosten und administrativen Kosten wie z.B. den Kosten für den Vertragsabschluss steht. Diese fallen gerade bei sehr geringen Stückzahlen ins Gewicht und machen kleinere Aufträge unrentabel. Dies ist nicht nur für das produzierende Unternehmen nachteilig, sondern auch für Auftraggeber/-innen, die ihre Aufträge nicht platzieren können.

Autonome Systeme können hier zwei Aspekte unterstützen. In der operativen Fertigung kann ein autonomes Fertigungssystem selbständig Modifikationen an Produkten vornehmen, um neue Produkte zu erzeugen. Ein Beispiel hierfür sind automatisierte Gussanlagen, die selbständig anhand von Spezifikationen Formen produzieren und dann automatisiert entsprechende Werkstücke gießen. Arbeiter/-innen müssen lediglich Verbrauchsmaterialien einfüllen und Spezifikationen laden. Zukünftig werden Autonome Systeme aber auch Verhandlungen führen können, um beispielsweise auf digitalen Marktplätzen Maschinenkapazitäten zu verkaufen und selbständig Aufträge und Lieferzeiten auszuhandeln, und sogar Preisverhandlungen werden sie führen können. Das wirtschaftliche Potenzial von Autonomen Systemen liegt in diesem Fall in der gesteigerten Effizienz und in der Fähigkeit, Abläufe wie die individualisierte Massenproduktion wirtschaftlich zu gestalten. Dies führt nicht nur zu einer kosteneffizienten Produktion, sondern auch zu einer deutlich breiteren Produktpalette, die wirtschaftlich produziert werden kann.

Autonome Systeme haben das Potenzial, Entscheidungswege zu verkürzen und verfügbare Ressourcen optimal zu nutzen. Die vorgestellten Beispiele unterstreichen die Fähigkeiten von Autonomen Systemen. Relativ monotone Aufgaben mit einem beschränkten Entscheidungsspielraum werden Autonome Systeme recht schnell übernehmen können. Komplexere Entscheidungen sind dann möglich, wenn die da-

für notwendigen Informationen in maschinenlesbarer Form bereitgestellt werden können. Hier gibt es laufende Forschungsvorhaben, die Informationen als Modelle oder Ontologien (vgl. Guarino/Oberle/Staab 2009) aufbereiten, um sie als Eingangsdaten auch für autonom agierende Systeme zu nutzen. Dies ist z.B. für die Automatisierung von Produktionsentscheidungen immens wichtig, um entscheiden zu können, welche Produktionsschritte sinnvoll von einer Maschine durchgeführt werden können.

Neben der wirtschaftlichen Produktion ist die wirtschaftliche Nutzung von Ressourcen ebenfalls ein Aspekt, den Autonome Systeme beeinflussen werden. Heute ist die Zeit, die Fahrzeugführer/-innen mit dem Fahren eines Kraftfahrzeugs verbringen, verlorene Zeit. Wenn diese Aufgabe von einem autonomen Fahrzeug übernommen werden kann, könnten sich die Passagier/-innen auf andere Tätigkeiten konzentrieren. Dies könnte die Vorbereitung des nächsten Termins, das Beantworten von E-Mails oder auch einfach die Entspannung nach einem anstrengenden Tag sein. Ebenfalls wird die Anzahl der benötigten Fahrzeuge im Zeitalter des autonomen Fahrens deutlich zurückgehen. Wenn Fahrzeuge autonom fahren können, ist es unsinnig, diese den größten Teil der Zeit auf Parkplätzen stehen zu lassen. Für viele Menschen wird die Nutzung von Mobilitätsdiensten gegenüber dem eigenen Fahrzeug die wirtschaftlichere Lösung sein. Ein solcher Dienst könnte so gestaltet sein, dass man kurzfristig per App ein Fahrzeug anfordern kann. Die Parkplatzsuche am Ziel entfällt und es ist immer dann ein Fahrzeug verfügbar, wenn es benötigt wird. Insgesamt werden damit deutlich weniger Fahrzeuge benötigt, um alle Mobilitätswünsche zu bedienen. Daher werden auch weniger Flächen für Parkplätze benötigt, die dann besser genutzt werden können, z.B. für Wohnungen oder Gewerbeflächen. Hier zeigt sich, dass Autonome Systeme auch die Nutzung von bestehenden und knapper werdenden Ressourcen wie Flächen in Städten positiv beeinflussen können und werden.

Die dargestellten Beispiele zeigen, dass die wirtschaftlichen Potenziale von Autonomen Systemen zunächst in deren Fähigkeit liegen, selbständig Entscheidungen zu treffen. Mittelbar ergeben sich daraus

allerdings eine Vielzahl von Potenzialen. Schnellere und optimierte Prozesse werden neue Geschäftsmodelle ermöglichen, die z.B. auf der Vermittlung von Ressourcen und deren effizienter Nutzung aufbauen. Teilweise haben Unternehmen dies schon erkannt und bauen Infrastrukturen auf, die dieses Konzept unterstützen. Man kann davon ausgehen, dass Angebote wie Uber, aber auch die zahlreichen Carsharing-Angebote, an denen Fahrzeugbauer beteiligt sind, sich in diese Richtung entwickeln werden und im Zeitalter der autonomen Fahrzeuge ihr Netzwerk nutzen werden, um Mobilitätsdienstleistungen anzubieten. Teilweise werden die wirtschaftlichen Potenziale von Autonomen Systemen auch eine Gefahr für etablierte Wirtschaftszweige darstellen. Werden im Zeitalter von autonomen Fahrzeugen weniger von diesen benötigt, dann werden Fahrzeughersteller andere Geschäftsmodelle entwickeln müssen, um ihre Einnahmen zu sichern. Gleiches gilt für die klassische Automatisierungstechnik, die im Zeitalter wandelbarer Fabriken und optimierter Auslastung weniger Geräte verkaufen wird. Die Geschäftsmodelle der Zukunft werden um digitale Ökosysteme herum realisiert werden, die einen effizienten Zugang zu Autonomen Systemen ermöglichen und dienstbasiert abrechnen. So wird z.B. nicht mehr das Fahrzeug, sondern die Mobilität verkauft. Es werden keine Produktionsanlagen oder Anlagenkapazitäten mehr verkauft, sondern eine gewisse Menge an Produkten und Werkstücken, die mit einer definierten Qualität gefertigt werden.

Gesellschaftliche Potenziale Autonomer Systeme

Neben ihren wirtschaftlichen Potenzialen haben Autonome Systeme auch signifikante gesellschaftliche Potenziale. Ein großes gesellschaftliches Problem ist heute die Landflucht. Umfragen zeigen, dass viele Familien sich ein Leben auf dem Land wünschen würden, jedoch die bessere Infrastruktur der Stadt bevorzugen. Kürzere Wege, ein besseres öffentliches Transportsystem, Einkaufsmöglichkeiten, öffentliche Infrastruktur und die Verfügbarkeit von Arbeitsplätzen sind die Hauptgründe, weswegen Menschen in die Städte ziehen. Die damit einherge-

hende Hektik, Lärmbelästigung, Stress und Enge werden in Kauf genommen. Autonome Systeme können hier in einigen Bereichen die Verhältnisse zwischen städtischen und ländlichen Regionen angleichen.

Im Bereich des öffentlichen Personennahverkehrs sind ländliche Regionen gegenüber Städten klar im Nachteil. Busse verkehren, wenn überhaupt, nur in großen Zeitabständen. Dies macht sie für die potenziellen Nutzer/-innen unattraktiv. Daher ziehen Personen, die diese Option haben, das individuelle Kraftfahrzeug als Transportmittel vor. Ältere Personen, die diese Möglichkeit nicht oder nicht mehr haben, müssen mit einer deutlich eingeschränkten Mobilität zurechtkommen. Dies führt schon bei einfachen Aufgaben, wie etwa einem Arztbesuch, zu Herausforderungen, z.B., wenn der Weg zu einem Facharzt/einer Fachärztin weit ist oder wenn häufiges Umsteigen erforderlich ist, um das Ziel zu erreichen.

Aufgrund der verhältnismäßig dünnen Besiedelung wird der Individualverkehr in ländlichen Regionen auch in Zukunft eine zentrale Rolle spielen. Autonome Fahrzeuge werden auch Menschen ohne Führerschein oder Menschen, die aufgrund ihres Alters oder aufgrund von Einschränkungen kein Kraftfahrzeug führen können, eine ausreichende Mobilität ermöglichen. Diese Autonomen Systeme haben hier das Potenzial, sich durch die verbesserte Mobilität direkt positiv auf die Lebensqualität von Menschen in ländlichen Regionen auszuwirken.

Neben der Beförderung von Personen ist die Beförderung von Gütern in ländlichen Regionen ebenfalls eine Herausforderung. Während in Großstädten kurze Wege zwischen Lager und Ziel die Kosten senken, müssen in ländlichen Regionen häufig lange Wege zurückgelegt werden. Dies ist insbesondere dann teuer, wenn nur wenige Güter in eine Region gebracht werden müssen, die Empfänger/-innen einer Lieferung nicht angetroffen werden und eine Fahrt wiederholt werden muss. Autonome Systeme können auch hier unterstützen, indem sie im Vorfeld ermitteln, ob der Empfänger/die Empfängerin einer Lieferung zu Hause ist und so unnötige Fahrten vermeiden.

Zuletzt werden Autonome Systeme auch die Lebensqualität zu Hause steigern. Assistenzsysteme haben das Potenzial, den Alltag zu erleichtern, z.B. indem sie die unterschiedlichen Aufstehzeiten der Bewohner/-innen berücksichtigen und diese möglichst sanft und angepasst an deren Schlafrhythmus wecken. Autonome Systeme können die Weckzeit z.B. davon abhängig machen, wann eine Person aufstehen muss, und zu diesem Zweck auf deren Kalender zugreifen. Aber auch nach einem längeren Krankenhausaufenthalt kann ein Autonomes System Unterstützung leisten. Zum Beispiel kann es den Tagesablauf einer Person verfolgen und bei ungewöhnlichen Abweichungen, wie z.B. mangelnder Aktivität, zuerst nachfragen und dann Hilfe holen. Auch bei Unfällen im Haus könnte ein solches System zum Einsatz kommen und im Fall der Fälle Hilfe für einen Verunglückten/eine Verunglückte rufen.

Weitere Potenziale Autonomer Systeme

Neben wirtschaftlichen und gesellschaftlichen Potenzialen haben Autonome Systeme auch überall dort Anwendungspotenziale, wo Entscheidungen in menschenfeindlichen Umgebungen getroffen werden müssen. Dies kann z.B. beim Rückbau eines Kernkraftwerks sein, wo Abschirmung und Strahlung den Funkverkehr zwischen Bediener/-innen und Roboter beeinflussen. Ebenfalls gibt es im Bereich der Erforschung des Weltraums und im Bereich der Tiefseeforschung, das heißt überall dort, wo die Kommunikation mit menschlichen Bediener/-innen schwierig ist und lange dauert, Anwendungspotenziale für Autonome Systeme, die selbständig Entscheidungen treffen.

HERAUSFORDERUNGEN UND LÖSUNGSANSÄTZE

Bis die Potenziale von Autonomen Systemen vollständig genutzt werden können, müssen verschiedene Herausforderungen gemeistert werden. Im Folgenden werden wir einige dieser Herausforderungen beschreiben und aktuelle Forschungsfragestellungen skizzieren.

Herausforderungen im Zusammenhang mit Kooperationsmodellen

Ein traditionelles eingebettetes System wird entwickelt, mit anderen Systemen vernetzt, in Verkehr gebracht, und erhält danach gegebenenfalls noch sporadische Updates, z.B., um Fehler zu beheben oder neue Funktionen hinzuzufügen. Diese Updates werden ebenfalls entwickelt und in einer definierten Umgebung getestet, die den Anwendungsbereich des Systems widerspiegelt. Autonome Systeme passen sich hingegen permanent an ihre Umgebung an und bilden dynamisch Verbünde mit anderen Systemen. Ob sich ein System in einen Verbund integriert, entscheidet es zur Laufzeit. Es koppelt sich mit einem Verbund, wenn diese Kopplung die Zielerreichung unterstützt. Die Beteiligung an einem Verbund ist in der Regel nur vorübergehend. Im Gegensatz zu traditionellen eingebetteten Systemen kann diese Integration nicht vorab durch Entwickler/-innen mittels Integrationstests getestet werden. Durch die Fähigkeit zur dynamischen Kopplung entsteht eine Vielzahl von Systemkombinationen, die die Funktionen einzelner Systeme in unterschiedlicher Weise koppeln. Ein Beispiel hierfür ist eine Kolonne von autonomen Fahrzeugen, die sich auf der Autobahn selbstständig formt, und bei der jedes Fahrzeug so lange Mitglied bleibt, bis die eigene Ausfahrt erreicht wurde. Um diese Kopplungen abzusichern, müssen neue Möglichkeiten für automatisierte Integrationstests zur Laufzeit erforscht werden. Eine mögliche Lösung könnte in einem virtuellen Integrationsraum bestehen, der Funktionen dynamisch in einer Simulation der Umgebung miteinander integrieren lässt. Dieser In-

tegrationsraum überprüft automatisch definierte Eigenschaften, wie z.B. die Kontrollierbarkeit und das Einnehmen sicherer Zustände, bevor die echten Systeme miteinander gekoppelt werden. Eine solche Integration erfordert Simulationsmodelle mit definierten Ausführsementiken und definierten Schnittstellen. Das *Functional Mockup Interface* (FMI) (vgl. Blochwitz et al. 2011) ist ein Kandidat hierfür, der derzeit von einer Vielzahl von Herstellern von Simulationswerkzeugen unterstützt wird.

Herausforderungen bei der Kommunikation

Eine weitere zentrale Herausforderung für kommunizierende Einheiten ist das Vorhandensein einer gemeinsamen Sprache. Menschen, die gemeinsam eine Aufgabe lösen sollen, werden ohne eine geeignete Kommunikationsform größte Mühe haben, ihre Arbeiten auf das gemeinsame Ziel hin zu koordinieren. Autonome Systeme sind mit der gleichen Herausforderung konfrontiert. Heute gibt es eine Vielzahl von Kommunikationsprotokollen und Kommunikationsparadigmen, die eine Integration von Systemen aus unterschiedlichen Bereichen erschweren.

Die Maschinen-zu-Maschinen-Kommunikation ist eine besondere Herausforderung für Autonome Systeme. Diese Systeme müssen mit einer Vielzahl von anderen Systemen kommunizieren können, auch wenn diese zu deren Entwicklungszeit noch nicht bekannt waren. Daher ist es erforderlich, system- und herstellerübergreifende Kommunikationsstandards zu definieren, die diese Kommunikation ermöglichen. Es gibt verschiedene Alternativen innerhalb der bestehenden Systemklassen, um hier eine Standardisierung zu erreichen. Zu nennen sind in diesem Kontext die Bemühungen um standardisierte Fähigkeitenmodelle durch die Plattform Industrie 4.0¹ oder die Standardisierungsaktivitäten des Car2Car Communication Consortium². Beide Initiativen de-

1 <https://www.plattform-i40.de> vom 20.8.2018.

2 <https://www.car-2-car.org> vom 20.8.2018.

finieren Standards, die den Datenaustausch zwischen Systemen und damit eine Maschinen-zu-Maschinen-Kommunikation ermöglichen.

Ein Beispiel für ein System, das von der Maschinen-zu-Maschinen-Kommunikation profitiert, ist ein dezentrales Logistiksystem, das die Auslieferung von Paketen in ländlichen Regionen organisiert. Pakete werden mithilfe dieses Systems in freien Kapazitäten von Pendler/-innen oder von Linienbussen befördert, um für Speditionen und Paketdienste unwirtschaftliche Strecken zu bedienen. Die Integration neuer, autonom agierender Systeme wie z.B. Drohnen erfordert gemeinsame Schnittstellen. Die Funktionen des Liefersystems, die die Lieferung von Paketen definieren, müssen mit den Fähigkeiten der Drohne zum Transport leichter Lasten integriert werden. Diese Integration zur Laufzeit erfordert ein grundlegendes gemeinsames Verständnis von Fähigkeiten, die z.B. den Transport von Gütern oder die Fähigkeit zur Bewegung mit einer bestimmten Geschwindigkeit beschreiben. Die zentrale Herausforderung hier liegt in der Definition von Schnittstellen und darin, die für eine dynamische Integration geeigneten Abstraktionsebenen zu finden. In dem genannten Beispiel würde das Liefersystem die Notwendigkeit beschreiben, einen Gegenstand mit einem bestimmten Gewicht innerhalb einer vorgegebenen Zeit zu transportieren. Die Drohne würde zur Integration ihre Fähigkeit zur Bewegung und zum Transport leichter Güter beschreiben. Für diese Kopplung wären aufgabenorientierte Schnittstellen geeignet. Das Logistiksystem definiert die Aufgabe, die durchgeführt werden muss, z.B. den Transport eines Pakets mit einem definierten Gewicht zu einem definierten Ort innerhalb einer definierten Zeit. Integrierte Systeme kommunizieren die Parameter, innerhalb derer sie diese Aufgaben realisieren können, ohne dass hierfür die Details der Umsetzung kommuniziert werden müssen. Durch semantische Technologien wie z.B. Ontologien können sich Systeme über Fähigkeiten informieren und prüfen, ob eine geforderte Fähigkeit von den angebotenen Fähigkeiten eines anderen Systems realisiert wird. Da Ontologien erweitert werden können, ist dies auch dann möglich, wenn eine bestimmte Fähigkeit zur Entwicklungszeit eines Systems noch nicht zur Verfügung stand.

Es wäre jedoch zu kurz gegriffen, nur die Kommunikation und die ausgetauschten Daten zu standardisieren. Die komplexeren Herausforderungen bei der Kopplung von Autonomen Systemen liegen in der Integration von Diensten. Dabei wird ein Code von einem System auf ein anderes System geladen, um dieses mit neuen Funktionen auszustatten. Durch die Integration von neuen Diensten können Autonome Systeme ihre Fähigkeiten auch nach ihrer Fertigstellung erweitern. Im Kontext des Logistiksystems könnte dies der eilige Transport von Medikamenten sein. Dieser Dienst wird möglicherweise von einer dritten Partei realisiert, die diesen Dienst in Software definiert und für die Umsetzung auf die Fähigkeiten bereits integrierter Systeme zurückgreift. Die oben genannten Drohnen könnten z.B. eilige Lieferungen realisieren. Dafür muss ein neuer Code mit dem Transportsystem integriert werden. Die dadurch entstehenden Herausforderungen bezüglich der Sicherheit vor Angriffen liegen auf der Hand. Bevor dieses Szenario realisiert werden kann, sind sichere Ausführumgebungen notwendig, die den fremden Code isoliert ausführen, Fehlverhalten oder schädliches Verhalten erkennen und in einem solchen Fall automatisch wieder einen funktionierenden Zustand herstellen können.

Absicherung Autonomer Systeme

Heute steuern eingebettete Systeme eine Vielzahl von sicherheitskritischen Prozessen. Dies sind z.B. Fahrfunktionen wie das Elektronische Stabilitätsprogramm (ESP) oder Produktionssysteme, die sicherstellen, dass ein Roboter seine Bewegungen einstellt, wenn ein Mensch eine Produktionszelle betritt. Heute werden diese Systeme dadurch abgesichert, dass z.B. mittels Fehlerbäumen (vgl. Vesely et al. 1981) die Ursachen von Fehlern identifiziert werden und geeignete Strategien im System verankert werden, die Fehler vermeiden oder deren Auswirkungen begrenzen. Für Autonome Systeme lassen sich diese bewährten Techniken nur eingeschränkt einsetzen. Durch die Fähigkeit, selbstständig Entscheidungen zu treffen, wird die Anzahl der möglichen Verhalten und Fehlerquellen so groß, dass diese sich nicht mehr a priori er-

mitteln lassen. Deutlich wird dies bei der Nutzung von neuronalen Netzen, beispielsweise zur Mustererkennung oder zur Entscheidungsfindung. Diese Systeme lassen sich aufgrund ihrer mathematischen Komplexität nicht mit bestehenden Methoden absichern, da diese auf dem Verständnis des internen Zustands eines Systems beruhen und auf dem Wissen, wie das System in einem Zustand auf einen definierten Stimulus reagiert. Hier sind neue Technologien, wie z.B. Safety Cages erforderlich, die Systeme nicht aufgrund ihres internen Zustands, sondern aufgrund ihres nach außen sichtbaren Verhaltens absichern.

Bei der Kopplung von Systemen zeigen sich Parallelen zu den oben beschriebenen Integrationstests. Wenn sich Autonome Systeme zu Verbünden integrieren, müssen auch für die Absicherung dieser Verbünde von Systemen neue Wege gefunden werden, um einen sicheren Betrieb zu garantieren. Garantiert werden muss z.B. die Qualität von ausgetauschten Daten und deren Aktualität sowie die Fähigkeit, auf einen Befehl innerhalb eines bestimmten Zeitfensters zu reagieren. Der dynamische Austausch von Sicherheitszertifikaten zum Aushandeln von minimalen Zusicherungen (vgl. Schneider/Trapp 2013) ist hier einer der möglichen Wege, um eine sichere Integration von Systemen zur Laufzeit zu erreichen.

Security

Da Autonome Systeme auch auf Daten zugreifen, ist deren Schutz eine weitere Herausforderung. Insbesondere dann, wenn persönliche Daten von Benutzer/-innen genutzt werden, muss deren Schutz vor einer missbräuchlichen Nutzung garantiert werden. Hierfür ist es wichtig, dass Nutzer/-innen ihren Systemen vertrauen können und wissen, was mit ihren Daten geschieht und was die Auswirkungen einer Einwilligung zur Nutzung sind. Hierfür gibt es Ansätze zur Datennutzungskontrolle, die darauf abzielen, auch technisch nicht versierten Benutzer/-innen die Kontrolle über ihre Daten zu ermöglichen (vgl. Steinebach et al. 2016). Aktuelle Forschungsvorhaben untersuchen beispielsweise, wie die Datennutzung durch Autonome Systeme dargestellt werden

kann, sodass deren Nutzer/-innen sinnvoll über erlaubte und verbotene Nutzungsszenarien entscheiden können.

Geschäftsmodelle

Die Nutzung von Autonomen Systemen wird neue Geschäftsmodelle ermöglichen. Während z.B. Carsharing heute hauptsächlich von Personen genutzt wird, die normalerweise auf andere Verkehrsmittel zurückgreifen und nur in bestimmten Situationen ein Auto benötigen, wird mit der Verbreitung autonomer Fahrzeuge die gemeinsame Nutzung von Fahrzeugen sprunghaft ansteigen. Schon allein deshalb, weil es nicht wirtschaftlich ist, ein autonom fahrendes Fahrzeug für eine längere Zeitdauer zu parken, werden autonom fahrende Fahrzeuge von mehreren Nutzer/-innen geteilt werden. Im Bereich der Produktion können Autonome Systeme selbständig Verträge aushandeln und die Produktion verändern, und so die Auslastung und den Gewinn eines Unternehmens maximieren.

Die Veränderung bestehender Geschäftsmodelle erfordert insbesondere von den menschlichen Nutzer/-innen Offenheit. In der Produktion findet man heute z.B. noch immer die Einstellung, dass eine Maschine, die nicht ausgelastet ist, nicht für einen Konkurrenten produzieren soll, auch wenn dies für beide Parteien einen Vorteil gegenüber Dritten bedeuten würde. Autonome Systeme werden es ermöglichen, vorhandene Ressourcen deutlich flexibler einzusetzen, und diese auch einzusetzen, um die Ziele von mehreren Systemen zu erreichen. Das nutzbare Potenzial dieser Systeme kann nur dann erreicht werden, wenn man dies zulässt. Beginnt man jedoch erst dann mit der Erforschung von Geschäftsmodellen, wenn die dafür notwendigen Autonomen Systeme schon existieren, riskiert man, von anderen Playern abgehängt zu werden. Die größte Herausforderung ist hierbei die Überwindung der ‚German Angst‘, der Angst vor Veränderungen.

DIGITALE ZWILLINGE ALS WEGBEREITER FÜR AUTONOME SYSTEME

Eine zentrale Herausforderung für jedes Autonome System, das Entscheidungen trifft, ist es, die Auswirkungen dieser Entscheidungen in seinem Kontext korrekt vorherzusagen und zu bewerten. Dabei werden ‚digitale Zwillinge‘ (vgl. Kuhn 2017) eine wichtige Rolle spielen. Digitale Zwillinge sind digitale Repräsentanzen von Dingen aus der realen Welt, die die relevanten Eigenschaften und Verhaltensweisen ihres realen Gegenstücks beschreiben. Sie repräsentieren dabei sowohl physische Objekte als auch nicht-physische Dinge wie z.B. Softwarekomponenten. Um eine sinnvolle Vorhersage zu ermöglichen, muss sich ein digitaler Zwilling in einer definierten Situation immer so verhalten, wie es sein reales Gegenstück auch tun würde. Dies bedeutet insbesondere, dass er sein reales Gegenstück mit der notwendigen Genauigkeit abbilden muss. Für einfache Vorhersagen können hier sehr einfache Modelle genügen. Ein Logistiksystem benötigt vielleicht nur die Information, wie lange ein Transport dauert. Andere Systeme dagegen benötigen wesentlich detaillierte Informationen, die z.B. auch die Interaktion mit anderen Systemen umfassen.

Ein zentraler Aspekt von digitalen Zwillingen ist die Fähigkeit, verschiedene Aspekte eines Systems mit einer einheitlichen Schnittstelle zu beschreiben. Sie umfassen zum einen Daten, die von dem Objekt der realen Welt stammen, aber auch Algorithmen, die ihr Gegenstück aus der realen Welt und dessen Verhalten akkurat beschreiben. Häufig handelt es sich dabei um Simulationsmodelle, die z.B. funktionale oder physische Eigenschaften des digitalen Zwillings simulieren. Werden diese Simulationsmodelle mit realen Daten ausgeführt, dann verhält sich der digitale Zwilling idealerweise genauso wie sein reales Gegenstück. Dies schafft die Möglichkeit, reale Tests durch virtuell durchgeführte Erprobungen zu ersetzen. Und das ermöglicht, automatisch Verbünde von Autonomen Systemen virtuell zu integrieren, diese im Hinblick auf geforderte Eigenschaften und Verhaltensweisen zu testen und deren Verhalten im Fehlerfall abzusichern.

Technisch erfordert dies einen virtuellen Raum, der digitale Zwillinge von unterschiedlichen Systemen integriert und ausführt. Dafür müssen digitale Zwillinge gemeinsame Schnittstellen besitzen, die eine Integration ihrer Simulationsmodelle ermöglichen. Das *Functional Mockup Interface* (FMI) (vgl. Blochwitz et al. 2011) ist eine Technologie, die dies grundsätzlich ermöglicht. Es führt Modelle in diskreten Zeitschritten aus und definiert den Datenaustausch zwischen Modellen anhand von definierten Ports. Simulationsmodelle werden als *Functional Mockup Units* gekapselt, die Simulationsmodelle und gegebenenfalls schon numerische Gleichungslöser beinhalten. Optional ist eine Simulation mehrerer Modelle mit einem gemeinsamen Gleichungslöser möglich. Um die Genauigkeit der gekoppelten Simulation zu erhöhen, können numerische Ereignisse, wie z.B. Sprünge in Funktionswerten, kommuniziert werden, sodass die Simulation um diese Werte herum von dem Gleichungslöser mit einer höheren Genauigkeit berechnet werden kann.

Durch eine kombinierte Simulation lässt sich das Verhalten komplexer Systemverbünde vorhersagen. Ebenfalls kann die Auswirkung einer Entscheidung auf die Qualität der eigenen Zielerreichung und auf die Zielerreichung von anderen Systemen mittels einer Simulation bewertet werden. Da in einem Verbund Autonomer Systeme alle Systeme jederzeit unabhängig voneinander Entscheidungen treffen können, ist es wichtig, dass alle Systeme eines Verbunds digitale Zwillinge besitzen, die sich zur Laufzeit integrieren lassen. Erst dann können alle beteiligten Autonomen Systeme die Auswirkungen ihrer Entscheidungen auf alle anderen Systeme berechnen und so überprüfen und gegebenenfalls zwischen der Erreichung eigener Ziele und der Erreichung von übergeordneten Zielen abwägen.

Die Kapselung von Simulationsmodellen in kompilierten *Functional Mockup Units* gewährleistet ebenfalls einen gewissen Schutz des enthaltenen geistigen Eigentums. Dieser Schutz ist jedoch nicht allumfassend und kann umgangen werden, wenn ausreichend Aufwand investiert wird. Müssen Simulationsmodelle auf sensible Algorithmen zurückgreifen, muss daher auch bei der Verwendung von digitalen

Zwillingen deren Schutz gewährleistet sein. Virtuelle Integrationsräume müssen daher auf vertrauenswürdigen Servern ausgeführt werden, die einen Schutz der digitalen Zwillinge vor dem Diebstahl geistigen Eigentums gewährleisten.

Neben der technischen Integration von Simulationsmodellen muss bei der Integration von digitalen Zwillingen sichergestellt sein, dass die ausgetauschten Daten zueinander passen. Auch wenn mithilfe des FMI-Standards eine durchgängig integrierte Simulation realisiert werden kann, scheitert ein solches Unterfangen in der Praxis oft am notwendigen Datenaustausch zwischen den unterschiedlichen Simulationsmodellen. Digitale Zwillinge müssen daher sicherstellen, dass alle relevanten Daten in einheitlichen Formaten bereitgestellt werden und dass bereitgestellte Daten auch von anderen Zwillingen in der gleichen Art verstanden werden. Dafür muss die Bedeutung eines Datenwertes klar definiert sein. Dies erfordert ein semantisches Verständnis der Bedeutung aller Daten. Ontologien bieten hierfür eine Lösung. Sie erlauben es, die Bedeutung von Daten maschinenlesbar zu definieren und dynamisch auszuwerten. Auch hier gibt es unterschiedliche Aktivitäten, die Ontologien bereitstellen und pflegen. Genannt seien hier eCI@ss³ und das IEC Common Data Dictionary⁴, die beide den fehlerfreien Datenaustausch zwischen Systemen unterschiedlicher Hersteller ermöglichen sollen. Beide Ontologien charakterisieren zentrale Elemente und deren Eigenschaften. Wenn sich die von digitalen Zwillingen bereitgestellten Daten auf Elemente der Ontologien beziehen, kann eine Software feststellen, ob es sich dabei um das gleiche Element handelt oder nicht. Neben der Gleichheit lassen sich in Ontologien auch weitere Eigenschaften von Elementen festhalten.

Mit Ontologien lassen sich z.B. auch die Fähigkeiten von Autonomen Systemen beschreiben. Durch die Gleichheitsrelation können Fä-

3 eCI@ss Klassifikation und Produktbeschreibung: www.eclass.eu vom 20.8.2018.

4 IEC 61360 – Common Data Dictionary (CDD): <https://cdd.iec.ch/cdd/iec61360/iec61360.nsf/TreeFrameset?OpenFrameSe> vom 20.8.2018.

higkeiten aus unterschiedlichen Ontologien aufeinander abgebildet werden. Auf diese Weise kann man der Situation begegnen, dass zwei unterschiedliche Systeme auch unterschiedliche Ontologien für die Dienstbeschreibung nutzen können. Durch die Definition einer Abbildung kann dennoch ein gemeinsames Verständnis der angebotenen Fähigkeiten realisiert werden.

ZUSAMMENFASSUNG

Zusammenfassend kann man ausführen, dass Autonome Systeme mit ihrer Fähigkeit zur selbständigen Entscheidung und Kooperation den Menschen in verschiedenen Bereichen unterstützen oder sogar ablösen werden. Für die Realisierung vollständig Autonomer Systeme sind jedoch noch einige Herausforderungen zu meistern, die Gegenstand aktueller Forschung sind. Diese umfassen einen einheitlichen Kommunikationsstandard; Fähigkeitenmodelle und Ontologien, die Fähigkeiten und Schnittstellen beschreiben; die Möglichkeit, zur Laufzeit das Verhalten integrierter Systeme zu testen; die Fähigkeit, die Auswirkungen von Entscheidungen auf das System und seinen Kontext vorherzusagen und die Möglichkeit, Autonome Systeme und Verbünde aus diesen zur Laufzeit abzusichern.

Diesen Herausforderungen stehen signifikante Potenziale Autonomer Systeme gegenüber. Autonome Systeme haben wirtschaftlich das Potenzial, bestehende Abläufe deutlich effizienter zu gestalten und z.B. die Produktion von kleinen Stückzahlen so zu gestalten, dass diese wirtschaftlich wird. Auch können sie Menschen von Routineaufgaben wie z.B. dem Autofahren entlasten, sodass diese etwas Produktiveres tun oder entspannen können. Gesellschaftlich können Autonome Systeme z.B. Mobilitätsdefizite beheben und es Menschen ermöglichen, länger in ihrer häuslichen Umgebung zu leben, ohne auf Pflege angewiesen zu sein. Ebenfalls lassen sich Autonome Systeme überall dort einsetzen, wo aufgrund lebensfeindlicher Bedingungen ein Einsatz von Menschen schwierig oder nicht möglich ist.

LITERATUR

- Blochwitz, Torsten/Otter, Martin/Arnold, Michael/Bausch, Constanze/Clauß, Christoph/Elmqvist, Hilding/Junghanns, Andreas/Mauss, Jakob/Monteiro, Manuel/Neidhold, Thomas/Neumerkel, Dietmar/Olsson, Hans/Peetz, Jörg-Volker/Wolf, Sebastian (2011): „The Functional Mockup Interface for Tool Independent Exchange of Simulation Models“, Konferenzbeitrag, Proceedings of the 8th Modelica Conference, 8th International Modelica Conference, Dresden, 20.-22.3.2011, <http://www.ep.liu.se/ecp/063/013/ecp11063013.pdf> vom 20.8.2018.
- Guarino, Nicola/Oberle, Daniel/Staab, Steffen (2009): „What is an Ontology?“, in: Steffen Staab/Rudi Studer (Hg.), Handbook on Ontologies, Berlin: Springer, S. 1-17.
- Kuhn, Thomas (2017): „Digitaler Zwilling“, in: Informatik-Spektrum 40/5, S. 440-444.
- Schneider, Daniel/Trapp, Mario (2013): Conditional Safety Certification of Open Adaptive Systems. ACM Transactions on Autonomous and Adaptive Systems 8/2, Artikel 8.
- Steinebach, Martin/Krempel, Erik/Jung, Christian/Hoffmann, Mario (2016): „Datenschutz und Datenanalyse. Herausforderungen und Lösungsansätze“, in: Datenschutz und Datensicherheit – DuD 40/7, S. 440-445.
- Vesely, William E./Goldberg, Francine F./Roberts, Norman H./Haasl, David F. (1981): Fault Tree Handbook. NUREG-0492, U.S. Nuclear Regulatory Commission, Washington DC: U.S. Nuclear Regulatory Commission.

Künstliche Intelligenz in Bildungs- und Arbeitsräumen

Internet der Dinge, Big Data,
Personalisierung und adaptives Lernen¹

Miloš Kravčik, Carsten Ullrich und Christoph Igel

1. EINLEITUNG

Die menschliche Entwicklung wird in hohem Maße von den verwendeten Werkzeugen und Medien beeinflusst. Die Forschung auf dem Gebiet der Neuronalen Plastizität hat gezeigt, dass unsere Erfahrungen unser Gehirn während des gesamten Lebens verändern (vgl. Doidge 2007). Menschen können also immer lernen und sich neues Wissen und neue Fähigkeiten aneignen. Gemeinsam mit anderen Informations-

¹ Wiederveröffentlichung des Beitrags mit freundlicher Genehmigung des Springer-Verlags, bei dem die Erstpublikation des Beitrags in englischer Sprache unter dem Titel "The Potential Internet of Things for Supporting Learning and Training in the Digital Age" in Zlatkin-Troitschanskaia/Wittum/Dengel (2018) erschien sowie mit freundlicher Genehmigung des Haufe-Verlags, bei dem der Beitrag unter dem Titel „Das Internet der Dinge als Lern- und Ausbildungshilfe in der digitalen Ära“ in Schwuchow/Gutmann (2018) in deutscher Sprache wiederveröffentlicht wurde.

und Kommunikationstechnologien (*Information and Communication Technologies* – ICT) hat das Internet einen gewaltigen Einfluss auf das Verhalten der Menschen (vgl. Carr 2011). Dabei kommt es auch zu ernststen Auswirkungen wie Verlust der Konzentrationsfähigkeit oder Probleme beim kritischen Denken und intensiven Lernen (vgl. Zlatkin-Troitschanskaia/Wittum/Dengel 2018). Jedes Medium entwickelt bestimmte kognitive Fähigkeiten und vernachlässigt andere (vgl. Greenfield 2009). Ob ein Werkzeug hilft oder schadet, hängt von den Fähigkeiten des Einzelnen und der Einsatzmethode ab. Darüber hinaus können sich die Langzeitauswirkungen diametral von den kurzfristigen Effekten unterscheiden und möglicherweise sehr schwer vorauszusagen sein. Es bestehen Risiken, die nicht ignoriert werden dürfen; wir müssen sie analysieren und versuchen, die sich daraus ergebenden Konsequenzen zu verstehen, um möglicherweise schädliche Auswirkungen auf den Menschen zu vermeiden. Dazu bedarf es einer Überwachung der langfristigen Auswirkungen der Technologie auf die Anwender/-innen und einer Abschätzung der möglichen Risiken (vgl. Zlatkin-Troitschanskaia/Wittum/Dengel 2018).

Zugleich sind die einzigartigen Möglichkeiten neuartiger ICT (z.B. das Internet der Dinge, *Wearables* und *Augmented Reality*) außerordentlich verlockend für Forscher/-innen und Entwickler/-innen auf dem Gebiet des technologiegestützten Lernens (*Technology Enhanced Learning* – TEL). Im Bildungsbereich kann uns eine richtige Analyse und Interpretation von *Big Data* dabei helfen, die Wirksamkeit und Effizienz von Lernprogrammen zu individualisieren und zu optimieren. Zusätzlich können die langfristigen Auswirkungen von ICT auf das Lernen überwacht werden, was auch bei der Kultivierung der so wichtigen metakognitiven Fähigkeiten wie Selbstkontrolle, Selbstüberwachung und Selbstreflexion hilfreich wäre. Diese Fähigkeiten sind unabdingbar beim Umgang mit den kognitiven Verzerrungen, die beim Menschen auftreten, und die häufig gegen uns eingesetzt werden. Die Nutzung dieser Verzerrungen zum Vorteil der Lernenden stellt aber auch eine große Herausforderung dar. Ein diesbezüglicher Erfolg wäre ein bedeutender Durchbruch auf dem Gebiet des TEL. In einer Zeit, in

der das lebenslange Lernen unverzichtbar geworden ist, spielt die Kultivierung von Lernfähigkeiten eine entscheidende Rolle. Lernen verlangt den Lernenden jedoch den Einsatz von Mühe und Energie ab. Die Stimulation der natürlichen Neugierde, der Umgang mit Unsicherheit sowie die Entwicklung der Kunst des Anzweifeln und des kritischen Denkens müssen daher Teil dieses Prozesses sein.

Ein weiterer wichtiger Aspekt ist die authentische Einbettung des Lernens in andere Prozesse wie die Arbeit. Dies hat direkten Einfluss auf die Motivation und somit auf die Effizienz des Lern- und Übungsprogramms, was nicht nur zur Aneignung von neuem Wissen, sondern auch zur Kultivierung bestimmter angestrebter Fähigkeiten führt. In einer Zeit, in der solch ein Paradigmenwechsel wie der Übergang zur Industrie 4.0 stattfindet, ist das von ausschlaggebender Bedeutung (vgl. Wahlster 2014).²

2. DAS INTERNET DER DINGE

Das Internet der Dinge (*Internet of Things* – IoT) ist eine „globale Infrastruktur für die Informationsgesellschaft, die erweiterte Dienstleistungen durch die Vernetzung (physischer und virtueller) Dinge auf der Grundlage existierender und entstehender dialogfähiger Informations- und Kommunikationstechnologien ermöglicht“ (ITU 2012: o.S., Übersetzung d.V.).

Das IoT kann die Authentizität von Lernprogrammen verbessern und die Motivation der Teilnehmer/-innen steigern, was einen ent-

2 Somit entstehen neue Herausforderungen für die berufliche Entwicklung, wobei Mitarbeiter/-innen direkt in die Planung und Umsetzung einbezogen werden. Lebenslange Kompetenzentwicklung ist aus unterschiedlichen Gründen wichtig. Unternehmen müssen ihre Kompetenzlücken erkennen und wirksam füllen. Mitarbeiter/-innen können ihre eigenen Ansprüche an ihre lebenslange berufliche Entwicklung haben und sie entsprechend planen.

scheidenden Einfluss auf die Wirksamkeit von Lernprozessen hat. Eine bessere Personalisierung und Anpassungsfähigkeit beim Lernen kann durch die gesammelten Informationen erreicht werden, die mittels einer breiten Palette an zur Verfügung stehenden Sensoren erfasst werden; dazu zählen beispielsweise Sensoren für Umweltanalyse, Gebäudeautomatisierung, Produktion sowie Bio-Sensoren. Dadurch entsteht potenziell eine neue Qualität des personalisierten Lernens auf der Grundlage eines besseren Verstehens der Anwender/-innen, ihrer aktuellen Verfassung (z.B. Aufmerksamkeit, Emotionen und Beeinträchtigungen) sowie ihres Kontextes. Es werden darüber hinaus neue Horizonte für Konzeption und Realisierung neuer virtueller Lernumgebungen eröffnet. In Kombination mit tragbarer Technologie (*Wearable Technologies* – WT) und erweiterter Realität (*Augmented Reality* – AR) wird es möglich, die menschlichen Sinne erheblich besser zum Lernen, zum Aneignen von Wissen und zum Trainieren neuer Fähigkeiten einzusetzen.

Aus technischer Perspektive besteht das IoT aus Objekten, die identifizierbar, kommunikations- und interaktionsfähig sind (vgl. Miorandi et al. 2012). Identifizierbar bedeutet, dass die Objekte über einen eindeutigen digitalen Bezeichner verfügen, den Elektronischen Produktcode (*Electronic Product Code* – EPC), der normalerweise mittels *Radio Frequency Identification* (RFID-Technologie), einer sehr einfachen Kommunikationsmethode, übertragen wird. Weitere Kommunikationsmöglichkeiten, z.B. Senden und Empfangen von Daten zwischen den Objekten, ergeben sich mittels unterschiedlicher Funktechnologien, womit der Schritt vom einzelnen Objekt zum Netzwerk der Dinge vollzogen wird. Die Objekte sind keineswegs passiv: Vielmehr setzen sie Sensoren zum Sammeln von Daten über ihre Umgebung ein. Akteure lösen dann entsprechende Aktionen aus. Mithilfe von Softwareebenen als Ergänzung der Hardware lassen sich auch Anwendungen ausführen. Die sogenannte *IoT-Middleware* wird normalerweise für den Zugriff auf heterogene IoT-Geräte genutzt und vereinfacht die Entwicklung von IoT-Anwendungen. Noch sind die technischen Herausforderungen des IoT nicht überwunden und seine unterschiedlichen

Teilbereiche sind Gegenstand aktiver Forschung. Dennoch ist die IoT-Technologie bereits so ausgereift, dass sie kommerziell eingesetzt und als Hilfsmittel in der Forschung und Bildung genutzt werden kann.

Frühe Arbeiten über den Einsatz des IoT zu Bildungszwecken konzentrieren sich hauptsächlich auf die Verwendung von RFID zur Erkennung eines Objekts und Darstellung einer Liste mit Datenelementen oder Aktivitäten für dieses Objekt (vgl. z.B. Broll et al. 2009). Später bestand dann auch die Möglichkeit, soziale Interaktionen in das Objektspektrum aufzunehmen (vgl. z.B. Yu et al. 2011). Die Forschung darüber, wie man das volle Potenzial des IoT für den Lern- und Bildungsbereich ausschöpfen kann, befindet sich noch immer in der Frühphase, wobei die bereits vorhandenen Arbeiten Herausforderungen und Chancen skizzieren und Architekturen beschreiben (vgl. Atif/Mathew/Lakas 2015; Thomas et al. 2012). Insbesondere die Forschung, die über die rein technologische Perspektive hinausgeht, ist noch äußerst lückenhaft.

3. BIG DATA IN DER BILDUNG

Zahlreiche Sensoren und von *E-Learning*-Plattformen erzeugte Protokolldateien tragen dazu bei, dass *Big Data* in der Bildung entstehen. Somit liegen riesige Datenmengen über die durchgeführten Lernprozesse vor. Die Analyse und das richtige Verstehen dieser Daten stellt eine echte Herausforderung für Forscher/-innen, Entwickler/-innen und auch für Anwender/-innen dar. Wie bereits erwähnt, ist der entscheidende Punkt die Lücke zwischen der technologischen und pädagogischen Perspektive (vgl. Zlatkin-Troitschanskaia/Wittum/Dengel 2018). Trotzdem helfen bestimmte Forschungsfelder dabei, diese Lücke zu schließen und ein Fundament zu schaffen, auf dem man aufbauen kann.

Educational Data Mining (EDM) zielt darauf ab, aus großen Lern-datenbeständen aussagekräftige Informationen automatisch zu extrahieren. Dabei wird auf maschinelle Methoden zum Erkennen von Da-

tenmustern zurückgegriffen. So lassen sich beispielsweise wirksame Lernpfade für einen bestimmten Anwender/eine bestimmte Anwenderin oder Aktivitäten, die zu besseren Benotungen führen, erkennen. Cristobal Romero und Sebastian Ventura (2010) führen die typischsten Aufgaben an, die in einer Lernumgebung durch EDM-Techniken bewerkstelligt wurden: Analyse und Visualisierung von Daten, Bereitstellung von Feedback zur Unterstützung von Lehrer/-innen bzw. Dozent/-innen, Empfehlungen und Vorhersage der Leistung von Schüler/-innen, Schülermodell-Erstellung, Erkennung unerwünschter Verhaltensweisen bei und Gruppierung von Schüler/-innen, Analyse des sozialen Netzwerks, Entwicklung von Konzeptkarten, Erstellung von Kursmaterial sowie Planung und Terminplanung.

Durch *Lernanalytik* (LA) sollen die Gesamteffektivität des Lernvorgangs verbessert und entsprechende Benachrichtigungen und Vorhersagen bereitgestellt werden. Die Verwendung dieser Methode lässt sich in fünf Stufen gliedern (vgl. Pardo 2014): Sammeln, Analysieren, Vorhersagen, Handeln und Einschränken. Auf jeden Fall müssen unterschiedliche Datenschutz- und Datensicherheitsebenen etabliert werden, damit, abhängig von den besonderen Präferenzen einzelner Personen und anderer Institutionen wie Unternehmen, verschiedene Integrationsniveaus möglich sind (vgl. Kravčík/Neulinger/Klamma 2016a). Die Visualisierung von beobachtbaren Daten über das Verhalten der Lernenden kann sogar ein Feedback über dessen nicht-beobachtbare kognitive und metakognitive Lernaktivitäten liefern (vgl. Nussbaumer et al. 2012). Ein wichtiger Anspruch, den man mit LA verbindet, ist die Messung des Lernerfolges nicht nur in formellen, sondern auch in informellen Umgebungen.

4. PERSONALISIERTES UND ADAPTIVES LERNEN

Wie bereits erwähnt, stellen *Big Data* in der Bildung eine wichtige Informationsquelle dar, die für das individualisierte Lernen zur Verfügung steht. Die Beschreibungen der Domäne können für einen bestimmten Anwender/eine bestimmte Anwenderin im aktuellen Kontext angepasst werden, und zwar je nach gewählter pädagogischer Methode und den bevorzugten Adaptionsstrategien. Um entsprechende technologische und konzeptuelle Differenzen zu überwinden, muss eine semantische Interoperabilität zwischen heterogenen Informationsressourcen und -diensten erreicht werden (vgl. Aroyo et al. 2006). Moderne kontextbezogene und lückenlose ICT macht *Smart Learning* möglich. Dadurch wird den Lernenden die richtige Hilfestellung angeboten, und zwar je nach aktuellem Kontext und personalisierten individuellen Bedürfnissen, die auch über das Verhalten der Lernenden bestimmt werden. *Smart Learning* setzt daher eine entsprechende Fusion aus Bildung und Technologie voraus (vgl. Li et al. 2016).

4.1 Benutzer- und Kontextmodellierung

Die Personalisierung eines Lernprogramms muss die Informationen über den Benutzer/die Benutzerin berücksichtigen, insbesondere das Lernziel (z.B. zu erwerbende Kompetenzen) sowie persönliche Präferenzen. Daraus ergibt sich dann die Wahl einer geeigneten pädagogischen Methode, um den Lernprozess wirkungsvoll, effektiv und attraktiv zu gestalten. Die tatsächlichen kontextbezogenen Einschränkungen führen zu Adaptionen, welche die aktuelle Umgebung sowie verfügbare Objekte berücksichtigen.

Olga Santos, Miloš Kravčik und Jesus Boticario (2016) benennen verschiedene Probleme bei der Personalisierung von Lernprogrammen, z.B. die effektive Erkennung und Verwaltung kontextbezogener und persönlicher Daten der Lernenden, wozu auch ihr emotionaler Zustand zählt. Das Ergebnis sollte ein besseres Verständnis der Person sein,

wozu Daten aus unterschiedlichen Ressourcen verarbeitet werden (z.B. tragbare Geräte mit physiologischen und Kontextsensoren) sowie entsprechende *Big Data*. Es ist ebenfalls wichtig, verschiedene Lernziele zu harmonisieren, z.B. kurz- und langfristige, je nach Präferenzen der Lernenden. Da sich diese Präferenzen schnell ändern können, spielen verfügbare Sensoren eine signifikante Rolle bei der Erkennung solcher Veränderungen. Mittels hochdimensionaler Daten, die von Sensoren erfasst werden, lassen sich auf der Grundlage des Verhaltens der Personen direkte Schlüsse auf die kontextbezogenen Präferenzen ziehen (vgl. Unger et al. 2017). Jedoch sollten die Lernenden und manchmal auch andere Beteiligte (z.B. Lehrer/-innen, Tutor/-innen, Eltern) wissen, welche Daten die Maschine über sie sammelt. Durch ein sogenanntes *Open Learner Model* wird die maschinelle Darstellung der Lernenden verfügbar und dient als wichtiges Mittel für die Lernunterstützung (vgl. Bull/Kay 2010).

Bisher fehlte die Unterstützung der Benutzermodellierung zur Verwertung und Verwaltung aus dem IoT gesammelter persönlicher Daten. Das Benutzermodellierungs-Framework IoTum (vgl. Kummerfeld/Kay 2017) soll diese Lücke füllen. IoT-Anwendungsentwickler/-innen fällt es damit leicht, schlanke, flexible, leistungsfähige und reaktionsschnelle Benutzermodellierungen vorzunehmen und einzusetzen, die zuverlässig, transparent und nachvollziehbar sind (vgl. Kay/Kummerfeld 2012). Andere Methoden befassen sich mit der Verarbeitung kognitiver Verhaltensmuster des Menschen (vgl. Raptis et al. 2017), emotionalen Zuständen (vgl. Sawyer et al. 2017) sowie mit der Modellierung psychomotorischer Aktivitäten (vgl. Santos/Eddy 2017).

4.2 Adaptives Lernen und Üben

Jede Form von Wissensvermittlung sollte auf einer fundierten pädagogischen Grundlage beruhen, die die Lernenden mit ihren Zielen, Fähigkeiten und Präferenzen, aber auch die Thematik und die kontextbezogenen Settings berücksichtigt. Ein geeignetes Anleitungskonzept stimuliert die Motivation der Teilnehmer/-innen und gestaltet die

Lernerfahrung gleichzeitig wirkungsvoll und effizient. Dennoch kann jeder Lernentwurf nur bis zu einem gewissen Grad spezifiziert werden; darüber hinaus muss er an konkrete Gegebenheiten, die sich mit der Zeit dynamisch ändern, angepasst werden. Diese Laufzeit-Adaption muss den Zustand und das Verhalten der Lernenden widerspiegeln und idealerweise den Flow-Zustand konservieren, der sich durch vollständige Aufmerksamkeit und Konzentration auszeichnet; dabei sind einerseits das Entstehen von Frust durch zu hohe Anforderungen und andererseits Langeweile aufgrund von trivialen Abläufen zu vermeiden.

Die Herausbildung von *metakognitiven Fähigkeiten* ist eine große Herausforderung, da diese direkten Einfluss auf die Lebensqualität einer Person haben und für das lebenslange Lernen unabdingbar sind. In diesem Zusammenhang lässt sich die Effektivität der Wissensvermittlung durch Anwendung des *selbstregulierten Lernens* (SRL) verbessern. Zu diesem Zweck müssen die eingesetzten Technologien einen individualisierten Ansatz sowie die richtige Balance zwischen der Freiheit der Lernenden und deren Führung unterstützen, um einerseits die Motivation zu stimulieren und gleichzeitig die Wirksamkeit und Effizienz des Lernprogramms zu berücksichtigen. Nützliche assistive Dienste müssen *offen* sein und die zur Verfügung stehende Technologie sowie die Lernkultur der Lernenden berücksichtigen. Gleichzeitig müssen sie *flexibel* sein und die richtige Mischung aus Adaptivität und Empfehlungen verfügbarer Optionen bereitstellen, um unterschiedlichste Ausprägungen von Führung/Anleitung und Freiheit zu ermöglichen (vgl. Nussbaumer et al. 2014). Zu einer wirkungsvollen SRL-Unterstützung gehört die Integration von Hinweisen in Form von personalisierten Info-Elementen (Benachrichtigungen, die mühelos umgangen werden können) und geeignete Reflexionsmöglichkeiten (vgl. Kravčík/Klamma 2014). In jeder Lernumgebung sorgen hilfreiche *Situationskenntnis-* und *Reflexionsdienste* für wertvolles Feedback für die Teilnehmer/-innen, die damit ihre metakognitiven Fähigkeiten festigen können. Die Schwierigkeit liegt in der sinnvollen Interpretation der während des Lernprozesses gesammelten Daten und in ihrer Darstellung in verständlicher Form. Damit dies durch den Einsatz nützlicher

Technologie gelingt, müssen Designer/-innen und Entwickler/-innen ihr Wissen aus unterschiedlichen Feldern, z.B. Bildung, Psychologie, Neurowissenschaft und Informatik einfließen lassen (vgl. Kravčík et al. 2017).

Das Einüben von Fähigkeiten nur anhand von Texten und Bildern gestaltet sich normalerweise schwierig. Deshalb sind hier dynamische und interaktivere Medien notwendig; sie stellen sogar eine Grundvoraussetzung für das Lernen im Beruf dar. Die Herausforderung beim Demonstrieren einer bestimmten Fähigkeit, Bedienung oder Handlung ist das Auffinden von relevanten Informationssegmenten aus einer riesigen Menge multimedialer Ressourcen für ein bestimmtes Objekt, einen bestimmten Kontext oder einen bestimmten Benutzer/eine bestimmte Benutzerin. Die Anwendung personalisierter und adaptiver Techniken auf mit Anmerkungen versehene Videodaten ist möglicherweise eine gute Richtung bei der Förderung des informellen Lernens am Arbeitsplatz (vgl. Kravčík et al. 2016b). Eine andere vielversprechende Alternative, die in Richtung *Smart Ambient Learning* vorstößt, bietet sich mit AR und WT (vgl. Koren/Klamma 2015). Ein automatisches Echtzeit-Feedback kann mit einer Kombination aus tragbaren Geräten, Stimmenanalyse und Bewegungssensoren bereitgestellt werden, z.B. beim Einüben nonverbaler Kommunikationsfähigkeiten für die freie Rede in der Öffentlichkeit (vgl. Schneider et al. 2016). Diese Ideen beschreiben neue Arten der Beurteilung, die auf der direkten Überwachung menschlicher Verhaltensweisen in authentischen Umgebungen basieren, wobei entweder ein Echtzeit-Feedback oder eine Leistungsanalyse für einen bestimmten Zeitraum bereitgestellt wird. Das Erlernen motorischer Fähigkeiten ist ein Bereich, in dem tragbare Technologie und Benutzermodellierung synergetisch miteinander kombiniert werden können, um unterstützend eingesetzt zu werden (vgl. Dias Pereira dos Santos/Yacef/Martinez-Maldonado 2017). Des Weiteren eröffnen sich neue Möglichkeiten für das verfahrensbezogene Training, z.B. Aufnahme und Wiedergabe der Performance eines Experten/einer Expertin, was immersives, situationsbezogenes und intuitives Lernen ermöglicht (vgl. Guest et al. 2017).

Eine entscheidende Einschränkung der verfügbaren Adaptionen- und Empfehlungsdienste liegt für gewöhnlich in ihrer fehlenden Durchschau- und Nachvollziehbarkeit, wobei es sich um ein typisches Problem handelt, wenn Techniken aus dem Bereich der Künstlichen Intelligenz wie ‚tiefes Lernen‘ eingesetzt werden (vgl. De Bra 2017). Im Rahmen der Anwendung zu Lernzwecken müssen solche maschinellen Entscheidungen anhand von Regeln oder Beweisen erklärbar sein, um das Vertrauen der Benutzer/-innen zu gewinnen. Grundsätzlich stimuliert ein Kontrollverlust negative Gefühle bei den Nutzer/-innen. Es sind deshalb außerdem klare und verwaltbare Datenschutzrichtlinien erforderlich (vgl. Golbeck 2017).

5. APPSIST-SYSTEM

Zur Demonstration einer intelligenten adaptiven Lerntechnologie, die den Weg zur Industrie 4.0 bereitet, führen wir das APPsist-System ein, welches die erste allgemein einsetzbare dienstorientierte Architektur mit unternehmensspezifischen Spezialisierungen darstellt. Die zugehörigen intelligenten Dienste umfassen benutzerzentrierte Unterstützung von Mitarbeiterqualifizierung und -schulung sowie benutzeradaptive kontextbezogene Unterstützung, die formalisiertes Expertenwissen nutzt.

APPsist ist ein Beispiel dafür, wie von Sensoren gesammelte Daten für die Aneignung von Wissen und für entsprechende Hilfestellungen eingesetzt werden. Ziel war die Entwicklung einer neuen Generation von mobilen, kontextabhängigen und intelligent-adaptiven Assistenzsystemen für die Wissensvermittlung und Maßnahmenunterstützung in der intelligenten Produktion. Die Forscher/-innen und Entwickler/-innen konzentrierten sich vorwiegend auf die Fähigkeiten und Kompetenzen der Mitarbeiter/-innen und auf Versuche, eventuell fehlende Fähigkeiten hinsichtlich der Erfüllung von Aufgaben am Arbeitsplatz zu kompensieren sowie auf die Maßnahmenunterstützung im Arbeitsprozess. Darüber hinaus fördern Dienste zur Wissensunterstützung die

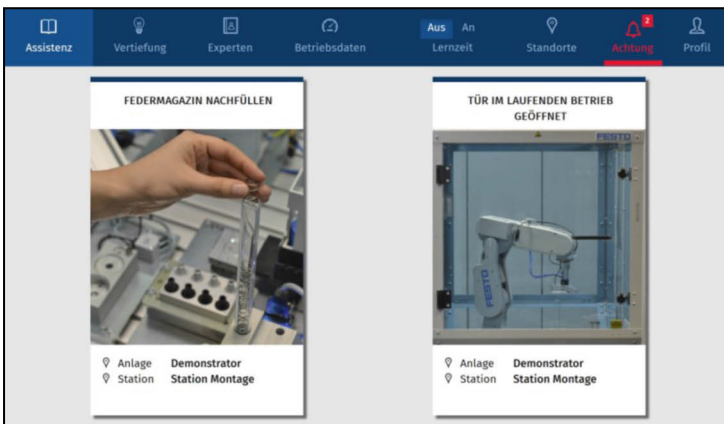
kontinuierliche Erweiterung von Mitarbeiterkompetenzen durch Aneignung von Wissen und Fähigkeiten in Bezug auf Fabrikation, Produkte und Prozesse. In dieser Hinsicht war das Ziel die berufliche Entwicklung der Mitarbeiter/-innen, damit sie schrittweise anspruchsvollere Aufgaben übernehmen und somit das Ungleichgewicht, das aufgrund des demographischen Wandels und des Fachkräftemangels entsteht, ausgeglichen wird. Dazu gehört die Einrichtung und der Betrieb einer Produktionseinheit im Produktionsprozess sowie die präventive Wartung, die reguläre Wartung und die Problembehandlung.

Die Lösung bietet den Mitarbeiter/-innen sowohl Assistenz- als auch Wissensvermittlungsdienste. Diese Softwarekomponenten bieten bestimmte Arten von Unterstützung: Assistenzdienste helfen bei der Lösung eines aktuellen Problems, während Wissensvermittlungsdienste der Übertragung von Wissen dienen; das heißt dem Erreichen individueller mittel- und langfristiger Entwicklungsziele (vgl. Ullrich et al. 2015). Eine solche Unterstützung bei einer bestimmten Aktivität kann in Form einer Schritt-für-Schritt-Anleitung oder als Einblendung von Informationen in das Sichtfeld der Nutzer/-innen mittels AR erfolgen (vgl. Abbildung 1). Kontextbezogene Empfehlungen enthalten geeignete Arbeitsschritte, aber auch Informationen, die im derzeitigen Kontext relevant sind, z.B. aus Handbüchern.

Der derzeitige Stand der Technik besteht in Dienstarchitekturen, deren Funktionalität sich aus dem Zusammenspiel einer großen Anzahl an Diensten ergibt. Jeder Dienst setzt dabei eine spezifische, unabhängige Funktionalität um und stellt diese anderen Diensten zur Verfügung. Das APPsist-System basiert auf einer dienstorientierten Architektur (vgl. Ullrich et al. 2015), die auf einen bestehenden Maschinenpark angewendet und damit verbunden werden kann. Es verwendet verfügbare Sensordaten, die dazu dienen, den Produktionsprozess zu überwachen und zu steuern, zum Zwecke der Interpretation der Handlungen der menschlichen Bediener/-innen, die mit den Maschinen interagieren. So kann das System den Bediener/-innen Vorschläge bezüglich durchzuführender Maßnahmen unterbreiten. Wenn APPsist beispielsweise einen Maschinenzustand erkennt, der einem Problem

entspricht, überprüft das System, welche Wartungsmaßnahmen das Problem lösen und welche Bediener/-innen befugt sind, diese Wartungsmaßnahme durchzuführen. Mittels einer mobilen Anwendung stellt es den Bediener/-innen dann zugehörige Inhalte (z.B. Bedienungsanleitungen, Hintergrundinformation) und Wartungsverfahren zur Verfügung. APPsist kann also personalisierte Lern- und Trainingsverfahren anbieten, die zum Aneignen des gewünschten Wissens oder der gewünschten Fähigkeit führen, indem es geeignete Arbeitsverfahren, aber auch entsprechende Lerninhalte empfiehlt. Diese Art von Unterstützung berücksichtigt die Entwicklungsziele der Arbeitnehmer/-innen sowie ihre durchgeführten Arbeitsschritte.

Abbildung 1: Der Screenshot zeigt empfohlene Arbeitsverfahren



Quelle: DFKI/EdTec Lab (2018)

Im Kontext der Verwendung des IoT zur Lern- und Trainingszwecken in der Produktion ist von Bedeutung, dass APPsist maschinelle Sensordaten mit Aktivitäten der menschlichen Bediener/-innen in Relation setzt und diese verwendet, um zu bestimmen, ob die Maßnahmen der Bediener/-innen richtig oder falsch waren (vgl. Ullrich et al. 2016). So werden in der ‚analogen‘ Welt durchgeführte Maßnahmen digital ver-

füßbar und zur Analyse, Interpretation und Reaktion nutzbar. Mit der fortschreitenden Digitalisierung von Räumen durch IoT-Technologie wird die Menge der für die digitale Verarbeitung zur Verfügung stehenden Daten weiter zunehmen. Es muss noch mehr Forschungsarbeit geleistet werden, um zu ergründen, wie solche Daten zum Lernen und Üben eingesetzt werden können; Beispiele wie APPsist zeigen, dass dies möglich ist.

6. PERSPEKTIVEN UND FAZIT

Heutzutage vermengt sich die Notwendigkeit der Weiterqualifikation aufgrund neuer Herausforderungen am Arbeitsplatz mit dem starken Bedürfnis, kritisches Denken zu fördern, um der systematischen Manipulation von Meinungen und mentalen Modellen, wozu die beim Menschen auftretenden kognitiven Verzerrungen ausgenutzt werden, entgegenzuwirken. Unter diesen Umständen ist das traditionelle Push-Model des in der Bildung üblichen formalen Lernens, das auf der Vermittlung von Wissen basiert, welches bereits vor der Ausführung des Lernvorgangs existierte, sehr häufig nicht ausreichend. Stattdessen bedarf es neuer Push-Modelle des informellen Lernens, wobei die Schaffung neuen Wissens während der Ausführung des Lernvorgangs gefördert wird (vgl. Naeve et al. 2008). Aus unterschiedlichen Perspektiven und Interpretationen eines bestimmten Ereignisses oder einer bestimmten Entität können die Lernenden ihr eigenes Verstehen schaffen, wobei ein gewisser Grad der Unsicherheit verbleibt. Bei diesem Vorgang spielt das Vertrauen in die Wissensressourcen, das ständig erneuert werden kann, eine wichtige Rolle. Eine der wichtigsten Kompetenzen wird die Fähigkeit sein, die richtige Frage zu stellen.

Die Industrie 4.0 bringt viele Herausforderungen mit sich und stellt hohe Ansprüche an das informelle Lernen, insbesondere direkt am Arbeitsplatz. Dadurch sollten sich neue Möglichkeiten für die Fort- und Weiterbildung von Mitarbeiter/-innen ergeben. Grundsätzlich sollte das Lernen nicht nur auf den individuellen Präferenzen der Benutzer/-in-

nen aufbauen, sondern auch auf der Wirksamkeit und Effizienz des Lern- und Übungsprogramms, wobei der aktuelle Kontext ebenso berücksichtigt werden muss wie der emotionale Zustand und die Aufmerksamkeit der Lernenden. Neue Sensoren und das IoT bieten weitere Alternativen für die Sammlung und Analyse von *Big Data* in formellen, informellen und arbeitsplatzgebundenen Lernprozessen. Durch sie lassen sich die Ziele, Präferenzen und der Kontext der Lernenden besser berücksichtigen, was zu einer präziseren Personalisierung und Adaption der Lernprogramme führt. Ihre Wirksamkeit und Effizienz lässt sich durch WT und AR weiter verbessern, wobei diese Technologien wahrscheinlich neue Trainingsmethoden zur Förderung der benötigten Kompetenzen hervorbringen. Ebenfalls von Interesse im Beruf und im Kontext des informellen Lernens ist eine Kombination aus *Learning Nuggets* und *Nudges*, das heißt die Erweiterung von *Micro-Learning*-Angeboten (*Nuggets*) mit passenden Empfehlungen (*Nudges*). Während sich Ersteres normalerweise auf kleinere Inhaltssegmente bezieht, kann Letzteres bei der Beförderung von Lernprozessen nützlich sein.

Das IoT ist dezentralisiert, da es autonome Geräte direkt miteinander verbindet. Verglichen mit herkömmlichen hierarchischen Modellen stellt es eine Alternative dar, die einen besseren Datenschutz und bessere Sicherheit bieten kann; Vertrauen ist hier der Schlüsselbegriff. Die *Blockchain*-Technologie (vgl. Tapscott/Tapscott 2016) ist entscheidend für das IoT, da sie den Geräten ermöglicht, digitale Verträge und Funktionen als selbstwartende Geräte autonom auszuführen. Dieses neue Paradigma delegiert das Vertrauen auf die Objektebene und ermöglicht die Animation und Personalisierung der physischen Welt. Des Weiteren schafft es neuartige ausgereifte Möglichkeiten für Benutzer/-innen, ihre Privatsphäre zu kontrollieren und ihre Daten zu schützen. Die *Blockchain*-Technologie hat das Potenzial, verschiedene Bereiche komplett zu verändern, und Bildung ist einer davon. Es gibt mehrere Möglichkeiten, wie dies geschehen kann (vgl. Tapscott/Tapscott 2017), und aus unserer Perspektive ist die wichtigste eine neue Bildung als Ersatz für das vorherrschende Sender-Modell und als Vorbereitung auf

das lebenslange Lernen. Dazu gehören alle Kompetenzen, die für Wissensarbeiter/-innen relevant sind, z.B. kritisches Denken, Problemlösung, Kollaboration und Kommunikation.

Das IoT kann bei der Sammlung riesiger Datenmengen über unterschiedliche Artefakte, ihre Beziehungen und die mit ihnen interagierenden Menschen einen bedeutenden Beitrag leisten. Die Schwierigkeit ist die richtige Analyse und Interpretation dieser Daten und die Gewinnung von aussagekräftigen Erkenntnissen über den Kontext und die einzelnen Benutzer/-innen. Auf dieser Grundlage lassen sich geeignete Lern- und Übungsprogramme entwerfen und den Menschen bereitstellen. Darüber hinaus schafft das IoT neue Möglichkeiten für die automatische Personalisierung und Anpassung dieser Programme an den aktuellen Kontext. So können Lernanweisungen sowie Personalisierungs- und Adaptionstrategien direkt in das IoT codiert werden, welches dabei als eine virtuelle (parallele) Maschine betrachtet werden kann. Dazu sind neue Programmiermethoden in diesem Bereich erforderlich. In jedem Fall müssen wir die Delegation von bestimmten Entscheidungsabläufen an Maschinen kontrollieren und überwachen, damit wir die langfristigen Auswirkungen beobachten und die verbundenen Risiken minimieren können.

LITERATUR

- Aroyo, Lora/Dolog, Peter/Houben, Geert-Jan/Kravčík, Miloš/Naeve, Ambjorn/Nilsson, Mikael/Wild, Fridolin (2006): „Interoperability in Personalized Adaptive Learning“, in: *Journal of Educational Technology & Society* 9/2, S. 4-18.
- Atif, Yacine/Mathew, Suijth S./Lakas, Abderahmane (2015): „Building a Smart Campus to Support Ubiquitous Learning“, in: *Journal of Ambient Intelligence and Humanized Computing* 6/2, S. 223-238.
- Broll, Gregor/Rukzio, Enrico/Paolucci, Massimo/Wagner, Matthias/Schmidt, Albrecht/Hussmann, Heinrich (2009): „PERCI: Pervasive

- Service Interaction with the Internet of Things“, in: IEEE Internet Computing 13/6, S. 74-81.
- Bull, Susan/Kay, Judy (2010): „Open Learner Models“, in: Roger Nkambou/Jacqueline Bourdeau/Riichero Mizoguchi (Hg.), *Advances in Intelligent Tutoring Systems*, Berlin: Springer, S. 301-322.
- Carr, Nicholas (2011): *The Shallows: What the Internet is Doing to Our Brains*, New York: Norton.
- De Bra, Paul (2017): „After Twenty-Five Years of User Modeling and Adaptation... What Makes us UMAP?“, in: Adán V. Sáenz/Newton N. Pereira/Luis M. Carral Couce/José A. Fraguera Formoso (Hg.), *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, New York: ACM, S. 1.
- Dias Pereira dos Santos, Augusto/Yacef, Kalina/Martinez-Maldonado, Roberto (2017): „Let’s Dance: How to Build a User Model for Dance Students Using Wearable Technology“, in: Adán V. Sáenz/Newton N. Pereira/Luis M. Carral Couce/José A. Fraguera Formoso (Hg.), *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, New York: ACM, S. 183-191.
- Doidge, Norman (2007): *The Brain that Changes Itself: Stories of Personal Triumph from the Frontiers of Brain Science*, New York: Penguin.
- Golbeck, Jennifer A. (2017): „I’ll be Watching You: Policing the Line between Personalization and Privacy“, in: Adán V. Sáenz/Newton N. Pereira/Luis M. Carral Couce/José A. Fraguera Formoso (Hg.), *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, New York: ACM, S. 2.
- Guest, Will/Wild, Fridolin/Vovk, Alla/Fominykh, Mikhail/Limbu, Bibek/Klemke, Roland/Sharma, Puneet/Karjalainen, Jaakko/Smith, Carl/Rasool, Jazz/Aswat, Soyeb/Helin, Kaj/Di Mitri, Daniele/Schneider, Jan (2017): „Affordances for Capturing and Re-enacting Expert Performance with Wearables“, in: Élise Lavoué/Hendrik Drachler/Katrien Verbert/Julien Broisin/Mar Pérez-Sanagustín (Hg.), *Proceedings of the 12th European Conference for Technology-Enhanced Learning (EC-TEL)*, Berlin: Springer, S. 403-409.

- ITU (2012): Internet of Things Global Standards Initiative. International Telecommunication Union, <http://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx> vom 19.5.2017.
- Kay, Judy/Kummerfeld, Bob (2012): „Creating Personalized Systems that People can Scrutinize and Control: Drivers, Principles and Experience“, in: ACM Transactions on Interactive Intelligent Systems (TiIS) 2/4, S. 24.
- Koren, István/Klamma, Ralf (2015): „Smart Ambient Learning with Physical Artifacts Using Wearable Technologies“, in: Davy Preuveneers (Hg.), Workshop Proceedings of the 11th International Conference on Intelligent Environments (= Ambivalent Intelligence and Smart Environments 19), Amsterdam: IOS Press, S. 325-332.
- Kravčík, Miloš/Klamma, Ralf (2014): „Self-Regulated Learning Nudges“, in: Mouzhi Ge/Francesco Ricci (Hg.), Proceedings of the First International Workshop on Decision Making and Recommender Systems (DMRS2014), Bozen, 18.-19.9.2014, CEUR 1278, S. 52-54, <http://ceur-ws.org/Vol-1278/> vom 20.8.2018.
- Kravčík, Miloš/Mikroyannidis, Alexander/Pammer-Schindler, Victoria/Prilla, Michael/Ullmann, Thomas D. (2017): „Editorial. In Awareness and Reflection in Technology Enhanced Learning [Special Issue]“, in: International Journal of Technology Enhanced Learning 9/2, S. 99-102.
- Kravčík, Miloš/Neulinger, Kateryna/Klamma, Ralf (2016a): „Data Analysis of Workplace Learning with BOOST“, in: Roberto Martinez-Maldonado/Davinia Hernandez-Leo (Hg.), Proceedings of the Workshop on Learning Analytics for Workplace and Professional Learning (LA for work). In Conjunction with the 6th International Learning Analytics and Knowledge Conference, Edinburgh, 25.-29.4.2016, CEUR 1601, S. 25-29, <http://ceur-ws.org/Vol-1601/> vom 20.8.2018.
- Kravčík, Miloš/Nicolaescu, Petru/Siddiqui, Aarij/Klamma, Ralf (2016b): „Adaptive Video Techniques for Informal Learning Support in Workplace Environments“, in: Ting-Ting Wu/Rosella Gen-

- nari/Yueh-Min Huang/Haroan Xie/Yiwei Cao (Hg.), *Emerging Technologies for Education. First International Symposium, SETE 2016, Held in Conjunction with ICWL 2016, Rom, 26.-29.10.2016* (= LNCS 10108), Cham: Springer, S. 533-543.
- Kummerfeld, Bob/Kay, Judy (2017): „User Modeling for the Internet of Things“, in: Adán V. Sáenz/Newton N. Pereira/Luis M. Carral Couce/José A. Fraguera Formoso (Hg.), *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, New York: AMC, S. 367-368.
- Li, Yanyan/Chang, Maiga/Kravčík, Miloš/Popescu, Elvira/Huang, Ronghuai/Niang-Shing Chen (2016), *State-of-the-Art and Future Directions of Smart Learning*, Singapur: Springer.
- Miorandi, Daniele/Sicari, Sabrina/Pellegrini, Francesco De/Chlamtac, Imrich (2012): „Internet of Things: Vision, Applications and Research Challenges“, in: *Ad Hoc Networks* 10/7, S. 1497-1516.
- Naeve, Ambjörn/Yli-Luoma, Pertti/Kravčík, Miloš/Lytras, Miltiadis D. (2008): „A Modelling Approach to Study Learning Processes with a Focus on Knowledge Creation“, in: *International Journal of Technology Enhanced Learning* 1/1-2, S. 1-34.
- Nussbaumer, Alexander/Scheffel, Maren/Niemann, Katja/Kravčík, Miloš/Albert, Dietrich (2012): „Detecting and Reflecting Learning Activities in Personal Learning Environments“, in: *Proceedings of the 2nd Workshop on Awareness and Reflection in Technology-Enhanced Learning (artel12) at European Conference for Technology-Enhanced Learning*, Saarbrücken, 18.9.2012, CEUR 931, S. 125-131, <http://ceur-ws.org/Vol-931/> vom 20.8.2018.
- Nussbaumer, Alexander/Kravčík, Miloš/Renzel, Dominik/Klamma, Ralf/Berthold, Marcel/Albert, Dietrich (2014): *A Framework for Facilitating Self-Regulation in Responsive Open Learning Environments*, <https://arxiv.org/ftp/arxiv/papers/1407/1407.5891.pdf> vom 20.8.2018.
- Pardo, Abelardo. (2014): „Designing Learning Analytics Experiences“, in: Johann A. Larusson/Brandon White (Hg.), *Learning Analytics. From Research to Practice*, New York: Springer, S. 15-38.

- Raptis, George. E./Katsini, Christina/Belk, Marios/Fidas, Christos/Samaras, George/Avouris, Nikolaos (2017): „Using Eye Gaze Data and Visual Activities to Infer Human Cognitive Styles: Method and Feasibility Studies“, in: Adán V. Sáenz/Newton N. Pereira/Luis M. Carral Couce/José A. Fraguela Formoso (Hg.), *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, New York: ACM, S. 164-173.
- Romero, Cristobal/Ventura, Sebastian (2010): „Educational Data Mining: A Review of the State of the Art“, in: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40/6, S. 601-618.
- Santos, Olga C./Eddy, Martha H. (2017): „Modeling Psychomotor Activity: Current Approaches and Open Issues“, in: Adán V. Sáenz/Newton N. Pereira/Luis M. Carral Couce/José A. Fraguela Formoso (Hg.), *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, New York: ACM, S. 305-310.
- Santos, Olga C/Kravčík, Miloš/Boticario, Jesus G. (2016): „Preface to Special Issue on User Modelling to Support Personalization in Enhanced Educational Settings“, in: *International Journal of Artificial Intelligence in Education* 26/3, S. 809-820.
- Sawyer, Robert/Smith, Andy/Rowe, Jonathan/Azevedo, Roger/Lester, James (2017): „Enhancing Student Models in Game-Based Learning with Facial Expression Recognition“, in: Adán V. Sáenz/Newton N. Pereira/Luis M. Carral Couce/José A. Fraguela Formoso (Hg.), *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, New York: ACM, S. 192-201.
- Schneider, Jan/Börner, Dirk/van Rosmalen, Peter/Speccht, Marcus (2016): „Can You Help Me with my Pitch? Studying a Tool for Real-Time Automated Feedback“, in: *IEEE Transactions on Learning Technologies* 9/4, S. 318-327.
- Schwuchow, Karlheinz/Gutmann, Joachim (Hg.) (2018): *HR-Trends 2019. Strategie, Digitalisierung, Diversität, Demographie*. Freiburg: Haufe.

- Tapscott, Don/Tapscott, Alex (2016): *Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World*, New York: Penguin.
- Tapscott, Don/Tapscott, Alex (2017): „The Blockchain Revolution & Higher Education“, in: *Educause Review* 52/2, S. 11-24.
- Thomas, Andrew M./Shah, Hanifa/Moore, Philip/Rayson, Peter/Wilcox, Anthony/Osman, Keith/Evans, Cain/Chapman, Craig/Athwal, Cham/While, David/Pham, Hai V./Mount, Sarah. (2012): „E-Education 3.0: Challenges and Opportunities for the Future of iCampuses“, in: *Proceedings of the Sixth International Conference on Complex, Intelligent, and Software Intensive Systems, CISIS 2012, Palermo, 4.-6.7.2012*, Piscataway: IEEE, S. 953-958.
- Ullrich, Carsten/Aust, Matthias/Dietrich, Michael/Herbig, Nico/Igel, Christoph/Kreggenfeld, Niklas/Prinz, Christopher/Raber, Frederic/Schwantzer, Simon/Sulzmann, Frank (2016): „APPsist Statusbericht: Realisierung einer Plattform für Assistenz- und Wissensdienste für die Industrie 4.0“, in: *Proceedings of DeLFI Workshop, CEUR 1669*, S. 174-180, <http://ceur-ws.org/Vol-1601/> vom 20.8.2018.
- Ullrich, Carsten/Aust, Matthias/Kreggenfeld, Niklas/Kahl, Denise/Prinz, Christopher/Schwantzer, Simon (2015): „Assistance- and Knowledge-Services for Smart Production“, in: *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*, New York: ACM, Artikel 40.
- Unger, Moshe/Shapira, Bracha/Rokach, Lior/Bar, Ariel (2017): „Inferring Contextual Preferences Using Deep Auto-Encoding“, in: Adán V. Sáenz/Newton N. Pereira/Luis M. Carral Couce/José A. Fraguella Formoso (Hg.), *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, New York: ACM, S. 221-229.
- Wahlster, Wolfgang (2014): „Semantic Technologies for Mass Customization“, in: Wolfgang Wahlster/Hans-Joachim Grallert/Stefan Wess/Hermann Friedrich/Tomas Widenka (Hg.), *Towards the In-*

ternet of Services. The THESEUS research program, Cham: Springer, S. 3-13.

- Yu, Zhiwen/Liang, Yunji/Xu, Bukan/Yang, Yue/Guo, B. (2011): „Towards a Smart Campus with Mobile Social Networking“, in: Feng Xia (Hg.), Proceedings of the International Conference on Internet of Things and 4th International Conference on Cyber, Physical and Social Computing, Dalian, 19.-22.10.2011, Piscataway: IEEE, S. 162-169.
- Zlatkin-Troitschanskaia, Olga/Wittum, Gabriel/Dengel, Andreas (Hg.) (2018): Positive Learning in the Age of Information. A Blessing or a Curse?, Wiesbaden: VS Verlag.

Grenzen und Widersprüche der Entwicklung und Anwendung ,Autonomer Systeme‘

Peter Brödner

1. EINFÜHRUNG: ÜBERZOGENE ERWARTUNGEN

Derzeit überrollt uns nach drei Jahrzehnten relativer Stille erneut eine Sturzflut von Meldungen über Projekte vermeintlicher ‚Künstlicher Intelligenz‘ (‚KI‘, engl. AI für *Artificial Intelligence*). Die Spannweite der Meldungen reicht von Heilsversprechen über Erfolgsgeschichten bis zu Szenarien der Apokalypse. Als kleine Auswahl aus dem gesamten Spektrum vermitteln die nachstehenden Äußerungen einen Eindruck:

- Unter der Überschrift „The future AI Company“ wird die künftige Automobilfabrik so beschrieben: „Superschlaue Computer, die ständig lernen, werden vieles übernehmen, was bisher Menschen erledigen: Sie antworten, wenn Kunden oder Lieferanten fragen, automatisch [...] [S]ie entwerfen sogar Autos und rechnen aus, wie sich die Entwürfe in der Fabrik umsetzen lassen.“ (Hofmann, VW-Vorstand, laut Schäfer 2016: o.S.)

- „Lernende Maschinen erkennen Gesichter und bringen sich selbst das Schachspielen bei.“ (Dworschak 2018: o.S.)
- „Invasion der Roboter: Künstliche Intelligenz ist bald so normal wie Strom.“ (Recke 2016: o.S.)
- Laut Google-CEO Pichai ist Künstliche Intelligenz für die Menschheit bedeutender als die Entdeckung des Feuers oder die Entwicklung der Elektrizität (vgl. Bastian 2018).
- Hawking „fears ‚AI may replace humans altogether‘ as a ‚new form of life that outperforms humans‘“. (Sulleyman 2017: o.S.)

Gegenüber solchen offenkundig sensationslüsternen, auf Verblüffung der Öffentlichkeit angelegten Äußerungen gilt es zunächst einmal den gesunden Menschenverstand zu bewahren und Aufklärung in wissenschaftlicher Analyse zu suchen. Auffällig an diesen und ähnlichen Äußerungen ist zunächst eine Gemeinsamkeit: Heilsverkünder/-innen wie Apokalyptiker/-innen sind gleichermaßen gegen Tatsachen immun. Gegen darin zum Ausdruck kommende ‚KI‘-Wahnvorstellungen hilft daher nur, die Wahrheit in relevanten Tatsachen zu suchen. Die Probleme beginnen freilich schon damit, dass es bis heute nicht gelingt, ‚KI‘-Systeme logisch zufriedenstellend von ‚gewöhnlichen‘ Computersystemen zu unterscheiden. Zwecks „Maschinisierung von Kopfarbeit“ (Nake 1992: 181) in Zeichenprozessen sozialer Praxis werden beide, auch gewöhnliche Computersysteme, immer schon zur Bewältigung von Aufgaben geschaffen „commonly thought to require intelligence“ (so eine übliche ‚KI‘-Definition, Autorengruppe 2018: 9). Dazu führen beide Systemarten gleichermaßen berechenbare Funktionen zur planvollen automatischen Verarbeitung zugehöriger Daten aus (vgl. Brödner 2008).

Bemerkenswert an Äußerungen wie den zitierten ist ferner, dass sie allesamt bestimmte Computerartefakte als solche in den Blick nehmen und diesen die Verblüffung hervorrufenden Eigenschaften zuschreiben. Dieser Fokussierung auf technische Artefakte liegt aber ein ebenso verbreitetes wie tief gehendes Missverständnis von Technik im Allgemeinen und der Computertechnik im Besonderen zugrunde. Einem

Bonmot des Philosophen José Ortega y Gasset (vgl. 1949) zufolge ist Technik ‚die Anstrengung, Anstrengungen zu ersparen‘. Damit trifft er den Kern der Sache: Genauere Analyse zeigt nämlich, dass Technik, verstanden als bloße Ansammlung technischer Artefakte, viel zu kurz greift. Artefakte fallen nicht vom Himmel, sondern müssen für bestimmte Zwecke mühsam konzeptionell entwickelt und materiell hergestellt werden. Als solche sind sie aber bloß tote, nutzlose Gegenstände, solange sie nicht für bestimmte Aufgaben zweckgemäß eingesetzt, mithin dafür angeeignet und praktisch wirksam verwendet werden. Das alles geschieht im Spannungsfeld des technisch Machbaren, der Formbarkeit von Natur, und des sozial Wünschenswerten, abhängig von jeweils herrschenden Interessen.

Nach allgemeinem professionellem Verständnis wird Technik daher definiert als die Gesamtheit von Maßnahmen zur Herstellung und zum Gebrauch künstlicher Mittel für gesellschaftliche Zwecke. Ihr werden damit nicht nur die Artefakte und Sachsysteme selbst zugerechnet, sondern gerade auch deren sozial konstruierte und kulturell vermittelte Herstellung und Anwendung (vgl. Ropohl 1991; Verein Deutscher Ingenieure 1991). Als geronnene Erfahrung verkörpern sie ein Stück sozialer Praxis und als Arbeitsmittel stellen sie Handlungsanforderungen an ihren Gebrauch, durch den Artefakte erst ihren Sinn erhalten und in ihrer Qualität zu beurteilen sind. Eben in den Prozessen der Entwicklung und Herstellung technischer Artefakte sowie ihrer Aneignung zu praktisch wirksamer Verwendung liegen gerade die eigentlichen Probleme der ‚Anstrengung, Anstrengungen zu ersparen‘; eben hierin liegen auch die Wurzeln missbräuchlichen Umgangs.

Sich in dieser Perspektive einigen der Probleme fortgeschrittener Computerentwicklung zu stellen, ist das Anliegen des vorliegenden Beitrags. Dazu werden im Folgenden zunächst am Beispiel sogenannter ‚künstlicher neuronaler Netze‘ (KNN) und Verfahren des ‚*Deep Learning*‘, die derzeit als vermeintliche Schlüsseltechnik der ‚KI‘ besonders hoch im Kurs stehen, deren Entwicklungsprobleme und Funktionsweisen aufgezeigt und hinsichtlich ihrer Tragweite beurteilt. Sodann werden die besonderen Einsatz- und Anwendungsprobleme be-

leuchtet, die mit diesen Systemen im Vergleich zu herkömmlichen Computersystemen verbunden sind. Vor dem Hintergrund eines kurzen historischen Exkurses über Meilensteine der Entwicklung von Computertechnik und *Computing Science* wird anschließend über verbreitete, tief sitzende Missverständnisse von Funktionsweise und Leistungsgrenzen von Computern aufgeklärt. Darauf fußend wird eine abschließende Bewertung vorgenommen.

2. ‚DEEP LEARNING‘ UND SEINE UNTERSCHÄTZTEN ENTWICKLUNGSPROBLEME

Neben Verfahren zur Analyse von ‚*Big Data*‘ auf Basis von Methoden schließender Statistik bieten sogenannte ‚künstliche neuronale Netze‘ (KNN) einen zweiten grundlegenden Ansatz der Verwirklichung sogenannten ‚maschinellen Lernens‘. Infolge exponentiell gesteigerter Leistung der Computer-Hardware sind sie in letzter Zeit zu einem bevorzugten Gegenstand der ‚KI‘-Forschung geworden. In Form vielschichtig strukturierter adaptiver Netzwerke bieten sie Verfahren zum sogenannten ‚*Deep Learning*‘ (vgl. LeCun/Bengio/Hinton 2015; Schmidhuber 2015) als einer Art neuer ‚Wunderwaffe‘.

Im Unterschied zu früheren ‚KI‘-Ansätzen der ‚allgemeinen Problemlösung‘ (vgl. Simon/Newell 1972) oder der ‚wissensbasierten Systeme‘ der ‚symbolischen KI‘ bzw. des ‚*Cognitive Computing*‘ (IBM), wie sie unter anderem beim ‚*Computer Integrated Manufacturing*‘ der 1980er Jahre verfolgt wurden, setzt die heute dominante neue Welle der ‚KI‘-Forschung nach einer langen Periode stark gebremster Aktivität nun vornehmlich auf die Entwicklung und den Einsatz von KNN als einem Grundmodell ‚lernfähiger‘ Systeme. Dieser Ansatz ist selbstverständlich keineswegs neu, sondern geht in seinen Anfängen zurück auf das ‚Perceptron‘ als einer biologisch inspirierten Nachahmung der logischen Funktionsweise von Nervenzellen bereits in der Zeit der allerersten materiellen Realisierungen von Computern (vgl. McCulloch/Pitts

1943). In Gestalt der ‚Lernmatrix‘ von Karl Steinbuch 1961 wieder aufgegriffen und zu größeren Netzwerken verknüpft (vgl. Hilberg 1995), fiel dieser Ansatz nach einem Verdikt durch Marvin Minsky und Seymour Papert (1969) erneut in einen Dornröschenschlaf, aus dem er in den 1990er Jahren langsam wieder erwachte. Wachgeküsst wurde er von der Aussicht auf ‚maschinelles Lernen‘ im Zusammenhang mit der Einsicht, dass die sich auf die logische Verarbeitung von durch Daten repräsentiertem explizitem Wissen stützende ‚symbolische KI‘ ihrerseits an den Grenzen der Explizierbarkeit von Können bzw. implizitem Wissen gescheitert war (vgl. Brödner 1997).

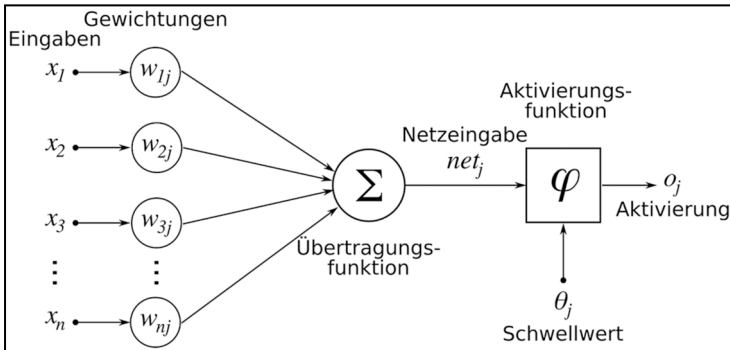
Versuche, kognitive Leistungen des Menschen, insbesondere sein intelligentes Handeln nachzuahmen, legen die Grundidee und den Ansatz nahe, das Gehirn mit seinen vielschichtig vernetzten Neuronen als Vorbild zu nehmen und einige seiner Strukturmerkmale und Funktionen möglichst direkt in Computersystemen nachzubilden. Freilich dürfen solche konnektionistischen Modelle nicht missverstanden werden als ‚naturgetreue‘ Nachbildungen des Gehirns oder des Zentralnervensystems, allenfalls gibt es gewisse strukturelle und funktionale Ähnlichkeiten, die durch den Aufbau des Gehirns inspiriert werden.

Ein KNN besteht im Wesentlichen aus einer Menge miteinander verbundener *Knoten*, die in Abhängigkeit von ihrem aktuellen Aktivierungszustand und der momentanen Eingabe ihren neuen Zustand bestimmen und eine Ausgabe produzieren. Diese Elemente sind gemäß der *Netzwerkstruktur* miteinander verknüpft. Diese kann als gerichteter gewichteter Graph oder durch eine *Konnektionsmatrix* dargestellt werden. Die Dynamik eines KNN wird beschrieben durch (vgl. Abbildung 1):

- eine Propagierungs- bzw. Übertragungsfunktion η , die aus den Ausgaben der vorgeschalteten Elemente sowie der Gewichtung der Verbindungen die aktuellen Eingaben in interne Netzwerkelemente berechnet,

- eine Aktivierungsfunktion σ , die für jedes Element dessen Aktivierung o_j als Ausgabe abhängig davon bestimmt, ob ein Schwellwert von der Netzeingabe net_j überschritten wird oder nicht.

Abbildung 1: Berechnungsfunktionen an einem Netzknoten



Quelle: Wikipedia, Chrislb, CC BY-SA 3.0, <https://commons.wikimedia.org./w/index.php?curid=224561> vom 20.8.2018

In der Regel besteht die Propagierungsfunktion aus einer einfachen Summenbildung der gewichteten Verbindungseinflüsse und die Aktivierungsfunktion wird meist für alle Elemente des Netzwerks einheitlich festgelegt. Innerhalb eines Netzwerks wird noch zwischen Eingabe-, Ausgabe- und internen Elementen unterschieden (*Hidden Units*, die zu tief gestaffelten *Hidden Layers* zusammengefasst werden, daher die Bezeichnung *Deep Learning*). Mittels verschiedener Typen von Ausgabe-, Propagierungs- und Aktivierungsfunktionen können Klassen von konnektionistischen Modellen gebildet werden.

Um ihre Aufgaben zu lösen, müssen KNN während des Entwurfs passend strukturiert und ihre Prozessoren mit einem bestimmten ‚Lern‘-Algorithmus gesteuert werden. Das Netz als Ganzes wird über diese Festlegungen hinaus nicht programmiert, sondern passt sich durch Veränderung der Gewichte nach Maßgabe eines den Nutzen maximierenden ‚Lern‘-Algorithmus an die spezielle Aufgabenstellung an

(daher die Benennung *adaptiv*). Beispielsweise werden bei Problemen der Musteridentifikation oder der Klassifikation – ein Aufgabentyp, bei dem künstliche neuronale Netze, vor allem solche vom Typ der ‚faltenden‘ oder ‚Convolutional Neural Networks‘ (CNN), besonders leistungsfähig sind – einer langen Reihe von Eingabemustern jeweils die zugehörigen Klassen als Ausgänge zugeordnet; aus diesen Zuordnungen vermag dann der Algorithmus mittels einer Nutzenfunktion automatisch passende Verbindungsgewichte zu bestimmen. Dies funktioniert auch bei Mustern, die durch explizite Merkmalsbeschreibungen schwierig oder gar nicht zu fassen sind (etwa bei der Identifikation handgeschriebener Buchstaben) – freilich mit Unsicherheiten. Meist ist dafür allerdings eine sehr große Zahl von Trainingsbeispielen (in der Größenordnung von 10^6) erforderlich (vgl. LeCun et al. 1998).

Bei der Anpassung ist für die Bestimmung der Gewichte $W_j := W_j - \eta \nabla_w L$ eines KNN die Berechnung des Gradienten $\nabla_w L(N(x))$ einer Nutzen- oder Verlustfunktion $L(N(x))$ erforderlich (mit $N(x)$ als Netzwerksausgabe, $L(N)$ als etwa über alle Trainingsbeispiele summierter euklidischer Distanz sowie der ‚Lernrate‘ η). Mit der Aktivierungsfunktion $\sigma(x)$ lässt sich – am einfachen Beispiel eines dreischichtigen KNN – die Funktionsweise des *Backpropagation*-Algorithmus und das häufig auftretende Problem des *schwindenden Gradienten* aufzeigen:

$$N(x) = W_1 \cdot \sigma \left(\overbrace{W_2 \cdot \sigma(W_3 \cdot x)}^{N_2} \right)_{N_3}$$

Die Bestimmung des Gradienten von L erfordert die Bildung der Ableitung von $N(x)$ als Verkettung mehrerer Funktionen nach der Kettenregel:

$$\begin{aligned} \frac{dL}{dW_1} &= \sigma(N_2) \cdot \frac{dL}{dN} \\ \frac{dL}{dW_2} &= \sigma(N_3) \frac{d\sigma(N_2)}{dN_2} \cdot W_1 \cdot \frac{dL}{dN} \\ \frac{dL}{dW_3} &= x \cdot \frac{d\sigma(N_3)}{dN_3} \cdot W_2 \cdot \frac{d\sigma(N_2)}{dN_2} \cdot W_1 \cdot \frac{dL}{dN} \end{aligned}$$

Dabei wiederholt auftretende Faktoren sind während des Trainings oft < 1 , sodass das Ergebnis infolge ihrer Multiplikation insgesamt gegen null tendiert – daher der ‚schwindende Gradient‘ und die Schwierigkeit, vielschichtige KNN zu trainieren. Zudem ist es schwierig, bei der Suche nach den Extremwerten der Nutzenfunktion eine variable, situativ passende Schrittweitensteuerung zu realisieren (vgl. Schmidhuber 2015; Wick 2017). Diese praktischen Schwierigkeiten bei der Strukturierung wie beim Training der Netzwerke sind allesamt nur durch jeweils fallspezifische Kunstgriffe zu überwinden, die das Können ihrer Entwickler/-innen herausfordern.

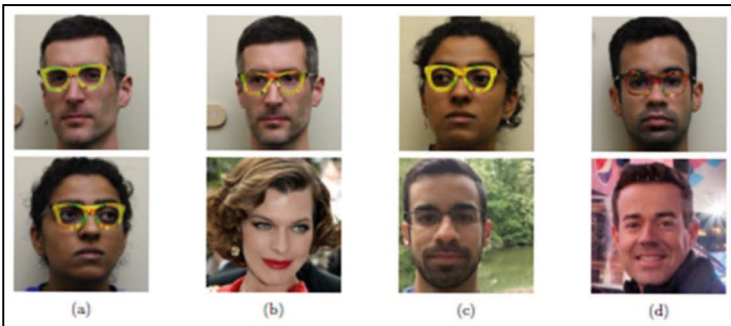
Für die Gestaltung der KNN gibt es keine theoretisch fundierten Erkenntnisse. Folglich müssen für jede Aufgabe passende Netzwerkstrukturen und Nutzenfunktionen mühsam mit großem Trainingsaufwand und ohne Erfolgsgarantie ausprobiert werden. Die Performanz der Netzwerke verdankt sich daher allein der Erfahrung, dem Können und der Kreativität ihrer Entwickler/-innen, darüber hinaus auch der in den letzten Jahren gemäß dem ‚Moore’schen Gesetz‘ enorm gesteigerten Leistungsfähigkeit von Computer-Hardware.

Zudem sind KNN in hohem Maße störanfällig. Schon durch geringfügige Veränderung der eingegebenen Bilddaten können sie in ihrer Funktionstüchtigkeit stark beeinträchtigt werden. So gibt es denn auch vielfältige Fehlleistungen von ansonsten erfolgreichen KNN und zahlreiche Beispiele sind belegt etwa bei der Bild-Klassifikation (vgl. z.B. Metzen et al. 2017; Nguyen/Yosinski/Clune 2015; Sharif et al. 2016; Sitawarin et al. 2016; Szegedy et al. 2014).

In Anbetracht der immer häufiger eingesetzten Systeme zur automatischen ‚Gesichtserkennung‘ ist das Beispiel eines mit einfachen Mitteln ausgetricksten Standardsystems zur Bildklassifikation nach neustem Stand der Technik besonders interessant und eindrucklich. Diese Systeme nutzen ebenfalls KNN, um im Falle von Gesichtsbildern anhand körperlicher Eigenheiten wie Position und Form der Nase oder Augenbrauen – mit Millionen Bildern trainiert – Personen voneinander zu unterscheiden. Werden diese Bereiche von einer Brille überdeckt, lässt das bunte Muster das KNN zur Gesichtsklassifikation Ei-

genheiten ausmachen, die fälschlicherweise als Gesichtsdetails ausgewertet werden. Ein männlicher Proband wurde so als die Schauspielerin Milla Jovovich erkannt (b), mit einer Genauigkeit von 87,9 Prozent; eine Asiatin hielt die Software mittels Brille für einen Mann aus dem arabischen Raum (c) etc. (vgl. Abbildung 2).

Abbildung 2: Ausgetrickste automatische ‚Gesichtserkennung‘¹



Quelle: Sharif et al. 2016: 1536

In letzter Zeit hat das System AlphaGo von Alphabet viel Aufsehen erregt, das in Turnieren die weltbesten Go-Spieler/-innen zu schlagen vermochte. Es wird häufig und zur Überraschung vieler als der ultimative Nachweis von dem Menschen überlegener ‚Künstlicher Intelligenz‘ und ‚maschinellem Lernen‘ präsentiert. Bei genauerem Hinsehen

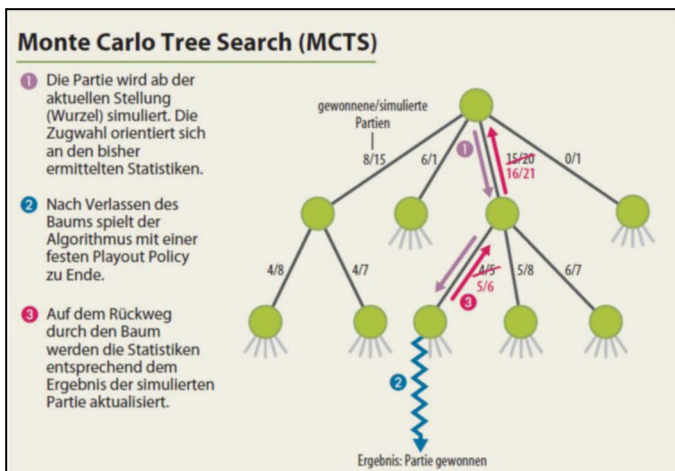
1 „Examples of successful impersonation and dodging attacks. Fig. (a) shows S_A (top) and S_B (bottom) dodging against DNN_B . Fig. (b)–(d) show impersonations. Impersonators carrying out the attack are shown in the top row and corresponding impersonation targets in the bottom row. Fig. (b) shows S_A impersonating Milla Jovovich (by Georges Biard / CC BY-SA / cropped from <https://goo.gl/GlsWIC>); (c) S_B impersonating S_C ; and (d) S_C impersonating Carson Daly (by Anthony Quintano / CC BY / cropped from <https://goo.gl/VfnDct>).“ (Sharif et al. 2016: 1536)

zeigt sich aber auch hier, dass diese Behauptungen auf dem propagandistischen Treibsand falscher Zuschreibungen gebaut sind.

Zunächst ist festzustellen, dass das Go-Spiel ein mathematisches Objekt ist, das durch seine Regeln vollständig definiert ist. Infolgedessen lässt sich jederzeit in jeder Stellung eines beliebigen Spielverlaufs eindeutig entscheiden, ob ein Spielzug erlaubt ist oder nicht. Im Prinzip ließe sich daher auch der Baum aller möglichen erlaubten Spielzüge und -verläufe darstellen (was freilich wegen sogenannter ‚kombinatorischer Explosion‘, hier der gigantischen Zahl von geschätzt rund 200¹⁵⁰ Zweigen, physisch unmöglich ist; es handelt sich um ein NP-vollständiges Problem).

Mittels heuristischer Verfahren muss daher die Suche nach erfolgreichen Spielzügen und einem möglichst optimalen Spielverlauf auf aussichtsreiche Teilbäume beschränkt werden. Als Methode bewährt hat sich bei vergleichbaren Aufgabenstellungen die Monte Carlo Tree Search (MCTS; vgl. Browne et al. 2012), die auch hier als ein heuristischer Verfahrens-Baustein zum Tragen kommt.

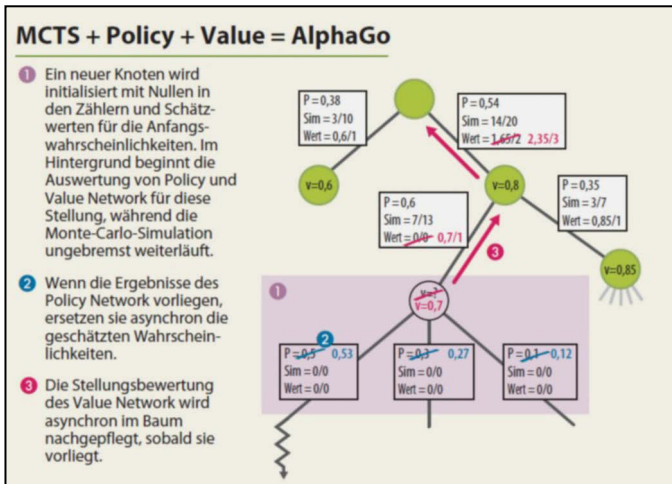
Abbildung 3: Funktionsweise der MCTS-Methode



Dabei werden für Folgezüge einer betrachteten Stellung – dargestellt als Wurzel des Teilbaums aussichtsreicher Züge – Erfolgshäufigkeiten (Verhältnis von Anzahl gewonnener zur Anzahl der insgesamt über diesen Zweig vollzogenen Spiele) mithilfe von parallel durchgängig simulierten Zufallspartien ermittelt, die laufend fortgeschrieben werden (vgl. Abbildung 3).

Im Falle von AlphaGo wurden dazu auf 40 parallel arbeitenden Prozessoren jeweils 10^3 Simulationen pro Sekunde durchgeführt (was bei einer Denkzeit von einer Minute zwischen den Zügen maximal 2,4 Millionen Simulationen pro Zug ermöglicht).

Abbildung 4: Kombination der MCTS mit KNN-basierten Bewertungen



Quelle: Bögeholz 2016: 151

In ihrer Leistung gesteigert wird die MCTS noch durch die Kombination mit zwei im Spiel gegen sich selbst trainierten neuronalen Netzen, die im Spielverlauf asynchron zusätzliche Bewertungen zur Zugwahl (*Policy*, in Form einer über die Zweige verteilten Erfolgswahrschein-

lichkeit) und zur Stellung (*Value*, als relativem Wert des zugehörigen Teilbaums) ermitteln. Dabei wächst der betrachtete Teilbaum aussichtsreicher Spielzüge durch Einführung neuer Knoten in besonders erfolgsversprechenden Zweigen mit anfangs geschätzten Bewertungsgrößen. Diese werden im Zuge der parallel und asynchron durchgeführten Läufe der MCTS-Simulationen und der Wertbestimmung für Zugwahl und Stellungen durch die neuronalen Netze fortgeschrieben, sobald sie verfügbar sind (vgl. Abbildung 4; weitergehende Einzelheiten der Verfahren und ihres dynamischen Zusammenspiels finden sich bei Silver et al. 2016 sowie Yuandong/Yan 2016).

Als mathematische, durch Regeln vollständig definierte Objekte sind Spiele entgegen landläufiger Auffassung geradezu prädestiniert für Modellierung und Formalisierung ihrer Spielverläufe. Es ist daher immer schon a priori sicher, dass es Algorithmen geben muss, die menschlichen Spieler/-innen überlegen sind. Erklärungsbedürftig ist folglich nur, warum sie erst jetzt gefunden wurden; Gründe dafür sind:

- Es bedarf der Entwicklung hinreichender methodischer Erfahrung und des notwendigen mathematischen Könnens, um leistungsfähige Heuristiken (z.B. MCTS) zur algorithmischen Bewältigung kombinatorischer Optimierung (oft NP-vollständiger Probleme) zu finden,
- es muss hinreichende Rechenleistung – etwa für das Training komplexer KNN – verfügbar sein.

Allgemein gilt für ‚KI‘-Systeme weiterhin: Grob irreführend als ‚künstlich intelligent‘ bezeichnete, *de facto* nur adaptive Computersysteme sind stets und ausschließlich das Ergebnis methodischer Kompetenz menschlicher Expert/-innen, deren Können und natürliche Intelligenz sie in Gestalt theoretischer Einsichten in zugrunde liegende Prozesse und raffiniert ausgeklügelter heuristischer Verfahren vergegenständlichen. Das gilt freilich für technische Artefakte, gleich welcher Komplexität, schon immer und manifestiert sich im Übrigen auch in der Berufsbezeichnung *Ingenieur* wie im Aristotelischen *téchne* als der Kunst, etwas beruhend auf Fachwissen, Übung und Erfahrung herstel-

len zu können. Zudem ist, wie gezeigt, das Verhalten von KNN als adaptiven Automaten durch deren Struktur und die Vorschriften des ‚Lern‘-Algorithmus zur passenden Veränderung der Gewichte vollständig determiniert (obgleich analytisch nicht mehr zu durchschauen).

3. EPISTEMISCHE UND ETHISCHE ANWENDUNGSPROBLEME VON ‚KI‘-ALGORITHMEN

Neben Herausforderungen und Schwierigkeiten der *Entwicklung* fortgeschrittener adaptiver Computersysteme bestehen aber auch auf Seiten der *Anwendung* große, wegen ihrer Besonderheit auch über den Einsatz herkömmlicher Systeme hinausweisende Probleme. Ihr Einsatz wirft gravierende ungelöste epistemische und ethische Fragen auf. Sich diesen Fragen zu stellen, ist umso dringlicher, je komplizierter und intransparenter Methoden der Modellierung von Zeichenprozessen und zugehörige algorithmische Verfahren werden und je leichter sie infolgedessen auch missbräuchlich verwendet oder manipuliert werden können (zu Details und Beispielen vgl. Autorengruppe 2018).

Infolge analytischer Intransparenz sind die Probleme adaptiver Systeme sowohl epistemischer wie ethischer Natur: Das Verhalten von ‚KI‘-Algorithmen (hier v.a. KNN und Verfahren schließender Statistik) ist selbst für Entwickler/-innen aktuell weder im Einzelnen durchschaubar („inconclusive evidence“) noch im Nachhinein erklärbar („inscrutable evidence“). Sie produzieren nur wahrscheinliche, daher stets unsichere Ergebnisse, deren Korrektheit und Validität nur schwer zu beurteilen sind. KNN sind zudem, wie oben beispielhaft gezeigt, sehr störanfällig und leicht auszutricksen. Die Ergebnisse, die sie liefern, sind in hohem Maße von der Qualität der Eingabedaten abhängig, die aber meist ebenfalls unbekannt oder nur schwer einschätzbar ist („misguided evidence“; Mittelstadt et al. 2016: 4f.).

Nutzer/-innen können dem Verhalten und seinen Ergebnissen daher nur blind vertrauen – trotz der nicht aufhebbaren Unsicherheit. Das

stellt sie in der ‚Koaktion‘ (siehe den Beitrag von Christoph Hubig in diesem Band, S. 289f.) mit solchen Systemen vor beträchtliche Belastungen: Wie sollen sie sich solche adaptiven Systeme überhaupt aneignen, wie mit ihnen zweckmäßig und zielgerichtet koagieren, wenn diese sich in vergleichbaren Situationen jeweils anders und unerwartet verhalten? Das wäre ein eklatanter Verstoß gegen eine der Grundregeln der Mensch-Maschine-Interaktion, gegen die Forderung nach erwartungskonformem Verhalten (vgl. EN ISO 9241-11-Anforderungen an die Gebrauchstauglichkeit). Zugleich würden auf Seiten der Nutzer/-innen stets aufs Neue überzogene Erwartungen an die vermeintliche Leistungsfähigkeit der Systeme geschürt, mithin gar ihre Wahrnehmung der Wirklichkeit verändert („transformative effects“, Mittelstadt et al. 2016: 5). Konfrontiert mit diesen Widersprüchen, unter dem Erwartungsdruck erfolgreicher Bewältigung ihrer Aufgaben einerseits und angesichts des Verlusts der Kontrolle über Arbeitsmittel mit undurchschaubarem Verhalten andererseits, würden sie unter dauerhaften psychischen Belastungen zu leiden haben (so bereits Norman 1994 im Hinblick auf ‚Agenten‘ als frühen adaptiven Systemen).

Darüber hinaus stellen sich mit diesen Systemen auch Fragen nach der ethischen Verantwortbarkeit: Dürfen derart undurchschaubare und störanfällige Artefakte überhaupt in die Welt gesetzt werden, da ihr künftiges Verhalten nicht sicher vorhersehbar ist? Wer ist gegebenenfalls für eingetretene Schäden verantwortlich – Entwickler/-innen, Betreiber/-innen oder Nutzer/-innen? – und wie werden daraus entstehende Haftungsansprüche geregelt? Bisher getroffene oder sich abzeichnende Regelungen sind unbefriedigend und unzureichend.

Auch hier gelten seit langem diagnostizierte ‚Ironien der Automatisierung‘ (vgl. Bainbridge 1983) zugespitzt weiter: Von besonderer Bedeutung ist dabei die Ironie, dass mangels hinreichender Übung und Erfahrung bei automatischem Normalbetrieb ausgerechnet die im Stör- oder Versagensfall wiederum benötigte menschliche Handlungskompetenz schwindet. Es fehlen Ansätze, wie dieser Art ‚erlernter Inkompetenz‘ entgegengewirkt werden kann. In Schadensfällen wird die Ursache meist dichotomisch in ‚menschlichem Versagen‘ oder einer techni-

schen Störung gesucht und dabei die wahre, in fehlgeleiteter sozio-technischer Systemgestaltung liegende Ursache ignoriert. Die derart systemisch bedingte mangelnde Kompetenz, in Verbindung mit der durch Untätigkeit oder Ablenkung geschwächten Vigilanz, führt dann bei plötzlich notwendigen Eingriffen zu beträchtlichen Problemen der Bewältigung:

- Unzureichende Erfahrungen und verlernte Fähigkeiten können zu fehlerhaften Diagnosen und falschen oder riskanten Aktionen führen,
- es entstehen, wie die Empirie ergibt, lange Verzögerungszeiten von sieben bis zehn Sekunden bis zum Zurechtfinden in der unerwarteten und ungewohnten Situation und zur Rückgewinnung von Handlungsfähigkeit (ein 150 km/h schnelles Auto fährt in dieser Zeit rund 400 m weit).

Beispiele und empirische Erkenntnisse zu diesen Herausforderungen gibt es aus der Forschung über Leitwartentätigkeiten, Flugführung oder automatisiertes Fahren zuhause, gleichwohl wurden bislang wenig weiterführende Konsequenzen gezogen. Daher ist infolge dieser noch weitgehend ungelösten Schwierigkeiten im praktischen Einsatz und Gebrauch adaptiver Systeme in naher Zukunft mit beträchtlichen Verzögerungen der Entwicklung zu rechnen (vgl. Bainbridge 1983; Baxter et al. 2012; Casner et al. 2014; Casner/Hutchins/Norman 2016; DIVSI 2016; Weyer 2007).

4. HISTORISCHER EXKURS: MEILENSTEINE DER ENTWICKLUNG VON COMPUTER- TECHNIK UND COMPUTING SCIENCE

Die hier angesprochenen Schwierigkeiten und Herausforderungen von Entwicklung und Anwendung fortgeschrittener adaptiver Computersysteme werden im landläufigen Verständnis von Computertechnik weitgehend ignoriert. Dadurch werden Fallstricke der Realisierung

verkannt, Potenziale maßlos übertrieben und gesellschaftliche Folgen falsch eingeschätzt. Dies ist aber beileibe keine neue Erscheinung, sondern begleitet die Computertechnik seit dem Beginn ihrer materiellen Manifestation, die gemeinhin mit rüstungstechnischen Entwicklungen im und kurz nach dem Zweiten Weltkrieg angesetzt werden. Seit her beherrschen prinzipiell irreführende, aber euphorisierende Metaphern wie ‚Elektronengehirn‘, ‚Künstliche Intelligenz‘, ‚maschinelles Lernen‘ oder ‚Autonome Systeme‘ bis hin zu den eingangs zitierten Äußerungen den gesellschaftlichen, oft aber auch den wissenschaftlichen Diskurs. Nur gelegentlich, in Zeiten großer Ernüchterung angesichts wirklicher Probleme der Modellierung und Algorithmisierung von Zeichenprozessen sozialer Praxis, ist weit zutreffender etwa von ‚elektronischer Datenverarbeitung‘ die Rede.

In der ansonsten irreführenden, durchweg anthropomorphisierenden Metaphorik kommt aber ein grundsätzlich fehlgeleitetes Verständnis von Computertechnik zum Ausdruck. Computersysteme führen, wie die Theorien der *Computing Science* lehren, mittels Daten als auf Syntax reduzierten Zeichen berechenbare Funktionen aus und sonst nichts. Ihr Verhalten ist durch die implementierten Algorithmen und eingegebenen Daten vollständig determiniert (nach dem Modell der Turingmaschine). Buchstäblich ‚wissen‘ sie nicht, was sie tun. Intelligent sind daher nicht die Computersysteme, sondern ausschließlich ihre Entwickler/-innen, die zuvor in mühevoller Arbeit mit ihrer Kreativität und methodischen Kompetenz Zeichenprozesse sozialer Praxis nach gewünschten Anforderungen modelliert, formalisiert und in berechenbare Funktionen überführt haben; oder ggf. auch deren Nutzer/-innen, die durch Aneignung ihrer Funktionen damit etwas Sinnvolles anzustellen vermögen (nähere Einzelheiten hierzu finden sich, gestützt auf den triadischen Zeichenbegriff von Peirce [1983 (1903)], bei Brödner [2008, 2018] sowie bei Nake [2001]). Diese professionelle Sicht wurzelt in der geschichtlichen Entwicklung von Computertechnik und *Computing Science* (die Bezeichnung ‚Informatik‘ ist ebenfalls eine irreführende Fehlbenennung: Es geht um Verarbeitung von Daten, nicht von ‚Information‘; vgl. Brödner 2014), wie sie in ihren logischen und

konzeptionellen historischen Meilensteinen zum Ausdruck kommt (vgl. Übersicht 1).

Tatsächlich beginnt die Entwicklung nicht erst im Zweiten Weltkrieg, sondern bereits in der Frühphase der industriellen Revolution mit der hochgradig arbeitsteiligen Organisation von Kopfarbeit, mittels derer anspruchsvolle kognitive Aufgaben (etwa die Neuberechnung umfangreicher mathematischer Tafeln im Dezimalsystem, z.B. Logarithmen, nautische Almanache, Artillerie-Schusstafeln) mittels Formularen in eine geplante Abfolge einfachster Rechenoperationen (Addition und Subtraktion reeller Zahlen mithilfe mechanischer Rechenmaschinen) aufgelöst und zur Ausführung vorgeschrieben werden. Dies geschieht in enger ideeller Verzahnung mit der arbeitsteiligen Organisation von Handarbeit in kleinste wiederkehrende Verrichtungen und ihrer anschließenden Mechanisierung mittels Arbeitsmaschinen (z.B. Textilmaschinen, Werkzeugmaschinen). Im Laufe der Entwicklung werden Funktionen des *Antriebs* („Kraftmaschinen“), der *Werkzeugführung* und der *Steuerung* zunehmend voneinander getrennt; dadurch werden für Antriebe fossile Energie (Dampfmaschinen mit Transmission, später Einzelantrieb mit Elektromotoren) nutzbar und Steuerungen realisiert, die statt Kräften Signale (Daten) über maschinelle Bewegungen verarbeiten, um deren gewünschten Ablauf zu gewährleisten. In dieser Perspektive lassen sich Computer auch als *universelles Steuerungspotenzial* begreifen, das per Programm in eine spezifische Steuerung verwandelt wird (wie es heute bei digitaler Prozesssteuerung üblich ist).

Die arbeitsteilige Organisation spezialisierter Verrichtungen erfordert zunehmend aufwendigere Planung und sachliche wie zeitliche Koordination der Einzelarbeiten durch „Manager/-innen“. Wachsender Aufwand für Koordination und Sicherung der Herrschaft über immer kompliziertere Prozesse der (Massen-)Produktion erfordern zudem Maßnahmen vereinfachender Standardisierung sowie wissensbasierter Planung, Anweisung und Kontrolle (Taylors Prinzipien des *„Scientific Management“*). Diese vertikale Arbeitsteilung der Trennung von Planung und Ausführung beruht auf expliziten Beschreibungen von Pro-

dukten und Prozessen, führt mithin zu einer ‚Verdoppelung‘ der Produktion in Zeichen (in Form von Zeichnungen, Stücklisten, Arbeitsplänen etc.). Das resultiert insgesamt in fortschreitender Verwissenschaftlichung von Produktion: Mit der Analyse, Planung und Kontrolle von Produktionsprozessen wird laufend erweitertes explizites Wissen über sie gewonnen und mit anderen wissenschaftlichen Erkenntnissen kombiniert. Dieses Wissen wächst wie ein Baum durch Verzweigung und wird durch Zeichen repräsentiert. So entstehen mit der ‚Verdoppelung‘ der Arbeitswelt in Zeichen auch zunehmend durch Zeichen repräsentierte Arbeitsgegenstände und ebensolche Methoden ihrer Verarbeitung. Die Anwendung dieses expliziten propositionalen Wissens zur Lösung wirklicher praktischer Probleme erfordert allerdings in wachsendem Maße wiederum Können, Wissensteilung und Kooperation von spezialisierten Wissensarbeiter/-innen.

Übersicht 1: Meilensteine der Entwicklung von Computertechnik und Computing Science

1792-1801:	<i>Gaspard de Prony</i> entwickelt ein formularbasiertes Verfahren zur extrem arbeitsteiligen Neuberechnung mathematischer Tafeln im Dezimalsystem; das Formular-Schema der Abfolge einfacher Rechenoperationen bildet die Urform eines Algorithmus (noch im Zweiten Weltkrieg wurden V2-Flugbahnen so berechnet).
1805:	Der <i>Jacquard-Webstuhl</i> war die erste digital mittels ‚Lochbrettern‘ gesteuerte Arbeitsmaschine.
1812:	<i>Charles Babbage</i> konzipiert die <i>Difference Engine</i> zur Berechnung der Funktionswerte von Polynomen: $f(x) = a_n x^n + \dots + a_1 x + a_0$ (teils öffentlich gefördert, 1822 realisiert).
Um 1830:	<i>Charles Babbage</i> entwirft und programmiert die allgemeinere verwendbare <i>Analytical Engine</i> . Sie nimmt die <i>Von-Neumann-Architektur</i> programmierbarer Uni-

versalrechner (Rechenwerk, Speicher, Steuerung, Dateneingabe und -ausgabe) vorweg, scheitert aber an der mechanischen Realisierung.

1847:	<i>George Boole</i> publiziert einen Logikkalkül (um 1888 von <i>Guisepe Peano</i> als <i>Boole'sche Algebra</i> axiomatisiert); er bildet das logisch-funktionale Fundament für <i>binäre Schaltsysteme</i> (als Kern heutiger Computer-Hardware).
1860-1880:	<i>Charles S. Peirce</i> entwickelt erstmals einen Prädikatenkalkül 1. Stufe, arbeitet an „logischen Maschinen“ und entwickelt die bislang elaborierteste <i>Theory of Signs</i> (triadische Zeichentheorie, ohne die computerisierte Kopfarbeit gar nicht zu verstehen wäre; äquivalent: <i>Gottlob Frege's</i> Begriffsschrift).
1931:	<i>Kurt Gödel</i> beweist die <i>Unvollständigkeit</i> formaler Systeme wie das der <i>Principia Mathematica</i> von <i>Bertrand Russell</i> und <i>Alfred North Whitehead</i> .
1937:	<i>Alan Turing</i> publiziert die Idee der <i>Turingmaschine</i> und definiert damit formal die Begriffe <i>Algorithmus</i> und <i>berechenbare Funktion</i> (äquivalent: <i>Lambda-Kalkül</i> von <i>Alonzo Church</i> und <i>Stephen Cole Kleene</i> 1936).
1945:	<i>Konrad Zuse</i> entwickelt den <i>Plankalkül</i> als erste algorithmische Programmiersprache (in Anlehnung an den <i>Lambda-Kalkül</i>).

Erst auf der Grundlage dieser Entwicklungsgeschichte der Computertechnik lässt sich ermesen, wie und warum Computer ihren Siegeszug durch die Arbeitswelt antreten konnten, sobald erst einmal mit elektromechanisch, später elektronisch realisierten binären Schaltsystemen die passende Hardware zur Verarbeitung binär codierter Daten als auf Syntax reduzierten Zeichen gefunden war. Im Zuge der ganzen Entwick-

lung bilden bis heute die logisch-konzeptionellen Ideen der Software mit den auf Erfahrung, Kreativität und Können ihrer Entwickler/-innen, auf deren 'lebendiges Arbeitsvermögen' (vgl. Pfeiffer 2004) angewiesenen Modellierungsmethoden und algorithmischen Verfahren den führenden Faktor und die Leistung der Hardware das limitierende Nadelöhr.

5. PRINZIPIELLE GRENZEN UND MISSVERSTÄNDNISSE DER COMPUTERTECHNIK

In diesem Zusammenhang ist zunächst an prinzipielle Grenzen der Formalisierung von Zeichenprozessen zu erinnern. Selbst die äußerst formalisierte Mathematik widersetzt sich ihrer vollständigen Algorithmisierung. Ausgerechnet im Zusammenhang mit den zu Beginn des 20. Jahrhunderts bestehenden großen Hoffnungen auf eine vollständige Formalisierung der bekannten Mathematik hat sich herausgestellt, dass es erwiesenermaßen unmöglich ist,

- einen Algorithmus anzugeben, der alle Sätze eines formalen Systems abzuleiten und deren Widerspruchsfreiheit zu zeigen imstande ist (vgl. Gödel 1931);
- einen Algorithmus anzugeben, der von jeder Formel eines formalen Systems entscheiden kann, ob diese Formel ein wahrer Satz des Systems ist (vgl. Turing 1965 [1937]).

Bezeichnenderweise beruht der Beweis von Gödel im Kern darauf, dass er als erfahrener und kompetenter Mathematiker eine Formel im System so zu konstruieren vermag, dass sie über einen durch ihn als wahr erkannten Satz aussagt, nicht beweisbar (ableitbar) zu sein. Zur mathematischen Fähigkeit von Menschen gehört eben auch, dass sie über alles, was sie mit deren Hilfe formalisieren können, durch Nach-

denken über die Formalisierung mittels abduktiven Schließens über sie hinaus zu gelangen vermögen (vgl. Brödner 1997, 2018).

Als rein formales Verfahren gibt die Abfolge von Operationen eines Algorithmus zwar Auskunft auf die Frage, was genau operativ abläuft; sie beantwortet aber nicht die Frage nach deren Sinn oder Bedeutung, warum sie so abläuft – eben deshalb gehört zur Software auch die Dokumentation mit derartigen Meta-Aussagen. Mit rein formalen Mitteln gelingt es eben nicht, auf der Metaebene Aussagen über Terme auf der operativen Ebene zu machen. Operationsfolgen sagen nichts über sich selbst aus, etwa ob sie korrekt oder gebrauchstauglich sind. So ist etwa auch die Frage, ob ein Algorithmus terminiert, formal nicht entscheidbar.

Das aus dem geschichtlichen Werdegang von Computertechnik und *Computing Science* gewonnene Verständnis der Funktionsweise von Computern als semiotische Maschinen erlaubt zudem, die fundamentalen Unterschiede zu Menschen als lebendigen Organismen aufzuzeigen (vgl. Übersicht 2). Die ständige Rede von ‚künstlich intelligenten‘ oder gar ‚autonomen‘ Computersystemen entpuppt sich dabei als folgenreicher Etikettenschwindel. Wieder einmal bedarf es der Philosophie als „Kampf gegen die Verhexung unseres Verstandes durch die Mittel unserer Sprache“ (Wittgenstein 1984: PU 109).

Übersicht 2: Ontologische Differenz zwischen Mensch und Computer

Mensch (lebendiger Organismus)	Computer (semiotische Maschine)
Sich durch <i>Autopoiese</i> in Stoffwechsel und Kommunikation <i>selber machend</i>	Wissensbasiert für bestimmte Zwecke <i>gemacht (konstruiert)</i>
<i>Autonom</i> (selbstbestimmte Regeln)	<i>Automatisch</i> (auto-operational, selbsttätig)
<i>Handelt intentional</i> (kontingent)	Verhält sich <i>kausal determiniert</i>

Ist sprachbegabt, <i>reflexiv lernfähig</i>	Ggf. algorithmisch gesteuert <i>Umwelt-adaptiv</i>
Lebendiges <i>Arbeitsvermögen</i> : <i>Können</i> (implizites Wissen, Erfahrung, situierte Urteilskraft & Handlungskompetenz), <i>verausgabt & reproduziert</i> sich im Gebrauch	Algorithmisch determiniertes <i>Verhalten</i> : Setzt <i>Formalisierung</i> von Zeichenprozessen voraus, muss für Praxis <i>angeeignet & organisatorisch eingebettet</i> werden

Die irreführende Metaphorik über Computer, als ob diese ‚wie Menschen‘ intentional eingestellt und handlungsfähig wären – ‚autonom‘, ‚selbstorganisiert‘, ‚intelligent‘, ‚smart‘, ‚selbstlernend‘, ‚selbstheilend‘ etc. –, ignoriert in reduktionistischer Weise nach Denkmustern des Funktionalismus die fundamentalen Unterschiede. Dadurch wird kompetentes Handeln von Menschen auf algorithmisch gesteuertes Verhalten von Maschinen reduziert; zugleich entstehen eben dadurch Illusionen über deren Zustandekommen und tatsächliche Leistungsfähigkeit. Das führt im Ergebnis zu einer verbreiteten Selbsttäuschung, wie sie in Diskursen um ‚Künstliche Intelligenz‘ zum Ausdruck kommt (vgl. Brödner 2018). In den Wahnvorstellungen vom vermeintlichen Eigenleben der Maschinen äußert sich deren Fetischcharakter, die „Macht der Machwerke über die Machenden“ (Haug 2005: 162).

6. FAZIT: GRENZEN UND WIDERSPRÜCHE ADAPTIVER SYSTEME

Aus diesen Ausführungen können mit Blick auf Grenzen und Widersprüche der Entwicklung und Anwendung fortgeschrittener adaptiver Computersysteme unmittelbar verschiedene Schlussfolgerungen gezogen werden:

Erstens bestimmen sogenannte ‚Autonome Systeme‘ die Regeln ihres Verhaltens nicht selbst, folglich sind sie tatsächlich nicht autonom, sondern als adaptive Automaten konstruiert.

Zweitens liegt es in der Regel-Natur von Spielen, ihrer Natur als mathematischem Objekt, dass Algorithmen – bei hinreichender Leistung der Hardware – menschlichen Spieler/-innen überlegen sind. Aus dieser Besonderheit kann aber nicht allgemein auf die Überlegenheit maschineller Verfahren über menschliche kognitive Kompetenz geschlossen werden.

Drittens gelingt es menschlicher Kreativität immer wieder, für spezielle, auch schwierige Aufgaben der Zeichenverarbeitung algorithmische Lösungsverfahren zu finden und als adaptive Automaten zu realisieren, die aber nur sehr begrenzt auf andere Aufgaben übertragbar sind. Meist ist die aufwendige Entwicklung jeweils eigener Methoden erforderlich, oft mit mehr Aufwand als Nutzen. Sogenannte ‚KI‘-Verfahren sind daher keine „General Purpose Technology“, wie das Erik Brynjolfsson, Daniel Rock und Chad Syverson (2017: 19f.) leichtfertig behaupten.

Viertens beruhen ironischerweise die Strukturen und Algorithmen erfolgreich eingesetzter adaptiver Systeme – mangels theoretischer Einsichten in Ursache-Wirkungs-Ketten – ausschließlich auf der Erfahrung und Kreativität, mithin dem Können bzw. dem Arbeitsvermögen ihrer Entwickler/-innen.

Fünftens ist das Verhalten adaptiver Systeme aus gleichem Grund nur schwer oder gar nicht zu durchschauen oder zu erklären, zudem äußerst störanfällig. In Form von KNN oder Verfahren schließender Statistik liefern sie stets nur wahrscheinliche, daher prinzipiell unsichere Ergebnisse. Das macht instrumentelles Handeln mit ihnen schwierig bis unmöglich, jedenfalls psychisch hoch belastend – gefordert sind daher Transparenz und Kontrolle ihrer algorithmischen Funktionen durch unabhängige Institutionen (wie bei anderer hochriskanter Technik auch).

Statt auf illusionäre ‚KI‘-Hoffnungen zu setzen, erscheint es angesichts dieser Einsichten in die zweifelhafte Tragfähigkeit von Konzepten ‚Künstlicher Intelligenz‘ weit aussichtsreicher, höhere Flexibilität, Produktivität und Innovationsfähigkeit stattdessen durch sozio-technische Gestaltung ‚guter Arbeit‘ zu erreichen. Dazu notwendiges Wis-

sen ist aus über drei Jahrzehnten Forschung zur Gestaltung von Arbeit und Technik verfügbar. Dies erscheint umso notwendiger, je mehr die künftige gesellschaftliche Entwicklung auf die breite Entfaltung menschlichen Arbeitsvermögens für den produktiven Umgang mit komplexen Beständen expliziten Wissens und technischen Systemen angewiesen ist. Gestützt auf dieses Wissen ließe sich das immer bedeutender werdende lebendige Arbeitsvermögen durch Organisation reflexiven und kreativen Zusammenwirkens kompetenter Expert/-innen mit gebrauchstauglich gestalteten Computersystemen weit wirksamer als durch den Einsatz von ‚KI‘-Systemen zur Entfaltung bringen.

LITERATUR

- Autorengruppe (2018): The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation, Oxford (AR): Future of Humanity Institute u.a. 02, <https://arxiv.org/pdf/1802.07228.pdf> vom 20.8.2018.
- Babbage, Charles (1971 [1832]): On the Economy of Machinery and Manufactures, New York: Kelley.
- Bainbridge, Lisanne (1983): „Ironies of Automation“, in: Automatica 19/6, S. 775-779.
- Bastian, Matthias (2018): „Google-Chef: Künstliche Intelligenz ‚wichtiger als Feuer und Elektrizität‘“, in: Vrodo vom 20.1.2018, <https://vrodo.de/google-chef-kuenstliche-intelligenz-wichtiger-als-feuer-und-elektrizitaet/> vom 21.8.2018.
- Baxter, Gordon/Rooksby, John/Wang, Yuanzhi/Khajeh-Hosseini, Ali (2012): „The Ironies of Automation ... still going strong at 30?“, in: Phil Turner/Susan Turner (Hg.), European Conference on Cognitive Ergonomics, ECCE '12, Edinburgh, 28.-31.8.2012, New York: ACM, S. 65-71, <https://dblp.uni-trier.de/db/conf/ecce/ecce2012.html> vom 21.8.2018.
- Bögeholz, Harald (2016): „Mysteriöse Tiefe. Wie Google-KI den Menschen im Go schlagen will“, in: c't 6, S. 148-151.

- Brödner, Peter (1997): *Der überlistete Odysseus. Über das zerrüttete Verhältnis von Menschen und Maschinen*, Berlin: edition sigma.
- Brödner, Peter (2008): „Das Elend computerunterstützter Organisationen“, in: Dorina Gumm/Monique Janneck/Roman Langer/Edouard J. Simon (Hg.), *Mensch – Technik – Ärger? Zur Beherrschbarkeit soziotechnischer Dynamik aus transdisziplinärer Sicht*, Münster: LIT, S. 39-60.
- Brödner, Peter (2014): „Durch ‚Information‘ desinformiert? Zur Kritik des Paradigmas der Informationsverarbeitung“, in: *Arbeits- und Industriesoziologische Studien* 7/1, S. 42-59.
- Brödner, Peter (2018): „Coping with Descartes’ Error in Information Systems“, in: *AI & Society Journal of Knowledge, Culture and Communication*, https://www.researchgate.net/profile/Peter_Broedner/publication/322553168_Coping_with_Descartes%27_error_in_information_systems/links/5a7ac73045851541ce5eff13/Coping-with-Descartes-error-in-information-systems.pdf?origin=publication_detail vom 23.8.2018.
- Browne, Cameron/Powley, Edward/Whitehouse, Daniel/Lucas, Simon/Cowling Peter I./Rohlfshagen, Philipp/Taverner, Stephen/Perez, Diego/Samothrakis, Spyridon/Colton, Simon (2012): „A Survey of Monte Carlo Tree Search Methods“, in: *IEEE Transactions on Computational Intelligence and AI in Games* 4/1, S. 1-49.
- Brynjolfsson, Erik/Rock, Daniel/Syversen, Chad (2017): *Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics*, NBER Working Paper 24001.
- Casner, Stephen M./Geven, Richard W./Recker, Matthias P./Schooler, Jonathan W. (2014): „The Retention of Manual Flying Skills in the Automated Cockpit“, in: *Human Factors* 56/8, S. 1506-1516.
- Casner, Stephen M./Hutchins, Edwin L./Norman, Don (2016): „The Challenges of Partially Automated Driving“, in: *CACM* 59(5), S. 70-77.
- Deutsches Institut für Vertrauen und Sicherheit im Internet (DIVSI) (2016): *Digitalisierte urbane Mobilität. Datengelenkter Verkehr zwischen Erwartung und Realität*, Hamburg: Deutsches Institut für

- Vertrauen und Sicherheit im Internet, <https://www.divsi.de/wp-content/uploads/2016/09/DIVSI-Studie-Digitalisierte-Urbane-Mobilitaet.pdf> vom 21.8.2018.
- Dworschak, Manfred (2018): „Was künstliche Intelligenz schon leisten kann – und was nicht“, in: Der Spiegel 2, <http://www.spiegel.de/spiegel/was-kuenstliche-intelligenz-schon-leisten-kann-und-was-nicht-a-1186438.html> vom 23.8.2018.
- Gödel, Kurt (1931): „Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I“, in: Monatshefte für Mathematik und Physik 38/1, S. 173-198.
- Haug, Wolfgang F. (2005): Vorlesungen zur Einführung ins „Kapital“, Hamburg: Argument.
- Hilberg, Wolfgang (1995): „Karl Steinbuch – ein zu Unrecht vergessener Pionier der künstlichen neuronalen Systeme“, in: Frequenz 49(1-2), S. 28-36.
- LeCun, Yann/Bengio, Yoshua/Hinton, Geoffrey (2015): „Deep Learning“, in: Nature 521, S. 436-444.
- LeCun, Yann/Bottou, Leon/Bengio, Yoshua/Haffner, Patrick (1998): „Gradient-Based Learning Applied to Document Recognition“, in: Proceedings of the IEEE 86/11, S. 2278-2324.
- McCulloch, Warren S./Pitts, Walter H. (1943): „A Logical Calculus of the Ideas Immanent in Nervous Activity“, in: Bulletin of Mathematical Biophysics 5/4, S. 115-133.
- Metzen, Jan H./Kumar, Mumtaz C./Brox, Thomas/Fischer, Volker (2017): Universal Adversarial Perturbations Against Semantic Image Segmentation, arXiv:1704.05712v1, <https://arxiv.org/pdf/1704.05712v1.pdf> vom 23.8.2018.
- Minsky, Marvin/Papert, Seymour (1969): Perceptrons, Cambridge, MA: MIT Press.
- Mittelstadt, Brent D./Allo, Patrick/Taddeo, Mariarosaria/Wachter, Sandra/Floridi, Luciano (2016): „The Ethics of Algorithms: Mapping the Debate“, in: Big Data & Society 3/2, S. 1-21.
- Nake, Frieder (2001): „Das algorithmische Zeichen“, in: Kurt Baudisch/Wilfried Brauer/Thomas A. Mück (Hg.): Informatik 2001,

- Tagungsband der GI/OCG Jahrestagung, Bd. II, Konstanz: UVK, S. 736-742.
- Nake, Frieder (1992): „Informatik und Maschinisierung von Kopfarbeit“, in: Wolfgang Coy/Frieder Nake/Jörg-Martin Pflüger/Arno Rolf/Jürgen Seetzen/Dirk Siefkes/Reiner Stransfeld (Hg.): *Sichtweisen der Informatik*, Braunschweig/Wiesbaden: Vieweg, S. 181-201.
- Nguyen, Anh/Yosinski, Jason/Clune, Jeff (2015): „Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images“, in: *Computer Vision and Pattern Recognition 15*, IEEE, <https://arxiv.org/pdf/1412.1897.pdf> vom 23.8.2018.
- Norman, Donald A. (1994): „How Might People Interact with Agents“, in: *CACM* 37/7, S. 68-71.
- Ortega y Gasset, José (1949): *Betrachtungen über die Technik*, Stuttgart: DVA.
- Peirce, Charles S. (1983 [1903]): *Phänomen und Logik der Zeichen*, Frankfurt a.M.: Suhrkamp.
- Pfeiffer, Sabine (2004): *Arbeitsvermögen. Ein Schlüssel zur Analyse (reflexiver) Informatisierung*, Wiesbaden: VS Verlag.
- Recke, Martin (2016): „Invasion der Roboter: Künstliche Intelligenz ist bald so normal wie Strom“, in: *t3n* vom 13.3.2016, <https://t3n.de/news/sxsw-traurige-roboter-688398/> vom 23.8.2018.
- Ropohl, Günter (1991): *Technologische Aufklärung. Beiträge zur Technikphilosophie*, Frankfurt a.M.: Suhrkamp.
- Schäfer, Ulrich (2016): „Leere Büros, leere Fabriken“, in: *Süddeutsche Zeitung* vom 10.11.2016, <http://www.sueddeutsche.de/wirtschaft/kuenstliche-intelligenz-leere-bueros-leere-fabriken-1.3243399> vom 23.8.2018.
- Schmidhuber, Jürgen (2015): „Deep Learning in Neural Networks. An Overview“, in: *Neural Networks* 61, S. 85-117.
- Sharif, Mahmood/Bhagavatula, Sruti/Bauer, Lujo/Reiter, Michael K. (2016): *Accessorize to a Crime. Real and Stealthy Attacks on State-of-the-Art Face Recognition*, in: *CCS '16 Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communica-*

- tions Security, <https://users.ece.cmu.edu/~mahmoods/publications/ccs16-adv-ml.pdf> vom 18.2.2018.
- Silver, David/Huang, Anja/Maddison, Chris J./Guez, Arthur/Sifre, Laurent/van den Driessche, George/Schrittwieser, Julian/Antonoglou, Ioannis/Panneershelvam, Veda/Lanctot, Marc/Dieleman, Sander/Grewe, Dominik/Nham, John/Kalchbrenner, Nal/Sutskever, Ilya/Leach, Madeleine/Kavukcuoglu, Koray/Graepel, Thore/Hassabis, Demis (2016): „Mastering the Game of Go with Deep Neural Networks and Tree Search“, in: *Nature* 529, S. 484-489.
- Simon, Hubert A./Newell, Allen (1972): *Human Problem Solving*, Englewood Cliffs, NJ: Prentice Hall.
- Sitawarin, Chawin/Bhagoji, Arjun N./Mosenia, Arsalan/Chiang, Mung/Mittal, Prateek (2016): DARTS: Deceiving Autonomous Cars with Toxic Signs, arXiv:1802.06430v1, <https://arxiv.org/pdf/1802.06430.pdf> vom 21.8.2018.
- Sulleyman, Aatif (2017): „Stephen Hawking Warns Artificial Intelligence ‚May Replace Humans Altogether‘“, in: *Independent* vom 2.11.2017, <https://www.independent.co.uk/life-style/gadgets-and-tech/news/stephen-hawking-artificial-intelligence-fears-ai-will-replace-humans-virus-life-a8034341.html> vom 23.8.2018.
- Szegedy, Christian/Zaremba, Wojciech/Sutskever, Ilya/Bruna, Joan/Erhan, Dumitru/Goodfellow, Ian/Fergus, Rob (2014): Intriguing Properties of Neural Networks, arXiv:1312.6199v4, <https://arxiv.org/pdf/1312.6199.pdf> vom 23.8.2018.
- Turing, Alan (1965 [1937]): „On Computable Numbers. With an Application to the Entscheidungsproblem“, in: Martin Davis (Hg.): *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvability Problems, and Computable Functions*, New York: Raven, S. 116-151.
- Verein Deutscher Ingenieure (Hg.) (1991): *Technikbewertung – Begriffe und Grundlagen. Erläuterungen und Hinweise zur VDI-Richtlinie 3780 (= VDI Report 15)*, Düsseldorf: VDI.

- Weyer, Johannes (2007): „Autonomie und Kontrolle. Arbeit in hybriden Systemen am Beispiel der Luftfahrt“, in: Technikfolgenabschätzung – Theorie und Praxis 16/2, S. 35-42.
- Wick, Christoph (2017): „Deep Learning“, in: Informatik Spektrum 40/1, S. 103-107.
- Wittgenstein, Ludwig (1984): Philosophische Untersuchungen, Werk-
ausgabe Bd. 1, Frankfurt a.M.: Suhrkamp, S. 225-580.
- Yuandong, Tian/Yan, Zhu (2016): Better Computer Go Player with
Neural Network and Long-term Prediction, arXiv:1511.06410v3,
<https://arxiv.org/pdf/1511.06410.pdf> vom 23.8.2018.

**Autonome Systeme
und der Wandel von Arbeit –
konzeptionelle Beiträge und
empirische Forschungsergebnisse**

Autonome Systeme und Künstliche Intelligenz

Herausforderungen für die Arbeitssystemgestaltung

Detlef Gerst

1. EINLEITUNG – INTELLIGENTE TECHNIK ALS HERAUSRAGENDES THEMA MODERNER ARBEITSGESTALTUNG

Heute meinen viele Menschen, moderne Produktionsmittel seien intelligent: Denn digitale Assistenten informieren eigenständig über Störungen und deren Ursachen. Regale bestellen Produkte nach, Kameras unterscheiden richtige von falschen Bauteilen und beurteilen, ob diese auch korrekt montiert sind. Menschen schreiben auch Produkten Intelligenz zu. Denn diese Produkte kommunizieren mit Maschinen, suchen sich eigenständig ihren Weg durch die Produktion und erinnern sich an den Prozess ihrer Entstehung.

Intelligente Technik ist ein herausragendes Thema in den Medien, der Politik, der Arbeitsforschung und auf Kongressen von Unternehmen und Arbeitgeberverbänden. Die Gründe hierfür sind verschieden. Unternehmensvertreter/-innen präsentieren sich so als modern. Für Politiker/-innen ist es Anlass, um dem Wahlvolk revolutionäre Veränderungen zu verkünden und ihre Programme zu rechtfertigen. Forscher/-innen prognostizieren Spektakuläres und können sicher sein, damit oft

zitiert zu werden und so an Profil zu gewinnen. Und Journalist/-innen berichten über neue, leistungsfähige Technik und verbinden das mit Ausblicken in eine beeindruckende Zukunft. Gemeinsam ist vielen, dass sie die Zukunft eher erschreckend denn erfreulich skizzieren.

Neue technische Produktionsmittel lösten auch in der Vergangenheit Ängste aus. Denn damals wie heute herrscht(e) das Argument, ein großer Teil der Menschheit verliere deshalb seine Beschäftigung oder die Kontrolle über die eigene Arbeits- und Lebenswelt. Die Aussicht, dass Technik eines Tages den Menschen ersetze oder beherrsche, begleitet die Modernisierung bis heute. Aber: Wirklich wurde diese Aussicht bis heute nicht. Vielmehr gilt: „Die Ideen, aus denen diese Prognose zusammengesetzt ist, sind fünfzig bis zweihundertfünfzig Jahre alt, und um ihre empirische Grundlage steht es heute nicht viel besser als früher.“ (Hirschi 2018: o.S.)

Heute sei jedoch alles anders, so die verbreitete Behauptung: Denn heute sei die Technik viel leistungsfähiger als früher, weshalb sich die Befürchtungen dieses Mal tatsächlich bewahrheiten werden. Als Beleg dient das Moor'sche Gesetz, wonach sich alle 18 Monate die Zahl der Transistoren auf einem Computerchip verdoppelt, während die Chips zugleich immer kleiner und billiger werden. Vergleichbar beeindruckend schnelle Innovationszyklen finden sich beispielsweise auch bei Sensoren, Speichern und Übertragungsraten von Daten. Erik Brynjolfsson und Andrew McAfee (vgl. 2014) prognostizieren deshalb, bald gebe es ‚brilliant technologies‘, welche die Arbeitswelt grundlegend verändern würden. Wenn Technik nicht heute schon Alleskönner sei, dann spätestens morgen oder übermorgen. Auch der enorme Energiebedarf, den die neuen Techniken benötigen, werde in der künftigen „Null-Grenzkosten Gesellschaft“ (Rifkin 2014) fast zum Nulltarif gedeckt.

Wichtig ist mir die Feststellung: Bei allen diesen Vorhersagen handelt es sich um Spekulationen. Richtig ist: Digitale Techniken verändern bereits heute die Arbeit in Produktion und Dienstleistung. Zunehmend werden sogenannte Autonome Systeme eingesetzt, die menschliche Arbeit ersetzen, ergänzen und in der Leistung oft übertreffen. Zunehmend arbeiten Menschen auch in unmittelbarer Interaktion

mit diesen Systemen. Autonome Systeme können mit Künstlicher Intelligenz (KI) verbunden werden, sie funktionieren aber auch ohne. Ich beschäftige mich mit den Fragen: Was bedeuten diese Systeme für die Arbeit von uns Menschen? Was ändert sich, wenn sie mit Künstlicher Intelligenz verbunden sind? Und wie können diese Systeme helfen, gute Arbeitsbedingungen zu schaffen?

2. BIG DATA, KÜNSTLICHE INTELLIGENZ UND AUTONOME SYSTEME

Computer und Roboter werden bereits seit Jahrzehnten eingesetzt, um Produktionsprozesse zu steuern, Produkte zu bearbeiten und administrative Aufgaben zu automatisieren. Heute ist, verbunden mit dem Begriff Industrie 4.0, von einer neuen Stufe der technischen Entwicklung die Rede. Grundlage sind sogenannte cyber-physische Systeme (vgl. Broy 2010). Der Begriff bezeichnet Produkte und Produktionsmittel, die es neben ihrer physischen Existenz noch einmal als Zwilling im digitalen Netzwerk gibt. Dieser digitale ‚Trick‘ befähigt Dinge, vermittelt über ihren Zwilling und mit Einsatz von Computerprogrammen, gemeinsam zu kommunizieren und zu entscheiden.

Dinge können vor allem mithilfe der KI gemeinsam handeln. Technik, die lernt und denkt, lässt sich jedoch auch in isolierten Systemen einsetzen. Wird automatisierte Technik mit KI ausgestattet, dann kann sie wesentlich mehr als nur starr programmierte Abläufe zu wiederholen. Technik kann nun lernen, die Veränderung von Rahmenbedingungen, auch wenn sie komplex sind, zu verarbeiten und die eigenen Aktivitäten anzupassen. Damit ist sie fähig, Aufgaben zu übernehmen, für die üblicherweise menschliche Intelligenz erforderlich ist. Das alles kann Technik, die mit KI ausgestattet ist: über Bestellungen entscheiden, Kapazitäten in Unternehmen planen, Personal auswählen, Lager verwalten, Produkte transportieren und verteilen, Wareneingang kontrollieren, Qualität prüfen.

Trotzdem: Die KI-Forschung ist auch heute noch weit von einer technischen Intelligenz entfernt, die der menschlichen gleicht. Der Begriff KI wurde erstmals im Jahr 1956 verwendet; als Titel einer Konferenz in Dartmouth, New Hampshire. Der Konferenz folgte eine Phase hochtrabender Erwartungen, die von zunehmender Ernüchterung abgelöst wurde. In den 1970er Jahren war die Enttäuschung über ausbleibende Erfolge so groß, dass die Forschung gravierend verringert und von einem ‚Winter der Künstlichen Intelligenz‘ gesprochen wurde. Heute ist KI dagegen wieder ein herausragender Gegenstand der Forschung. Immerhin sind anspruchsarme Formen von KI so weit entwickelt, dass sie in Produkten und Produktionsmitteln eingesetzt werden können und auch eingesetzt werden. Aus diesem Grund liegt heute die Zukunft der Arbeit, soweit sie von KI geprägt wird, vor allem in den Händen der Unternehmen.

In der öffentlichen Debatte spielen außer der KI auch die Begriffe Big Data, maschinelles Lernen und Autonome Systeme eine wichtige Rolle. Jeder dieser Begriffe steht für einen eigenen technischen Schwerpunkt, jedoch gibt es vielfache Überschneidungen und Wechselwirkungen.

2.1 Big Data, Data Mining und Smart Data – exponentielles Wachstum der Rechenleistungen

Big Data und Data Mining sind die Namen für Verfahren aus dem Bereich der *statistischen Datenanalyse*, mit denen große strukturierte Datenmengen gewonnen und statistisch analysiert werden können. Daten werden mit dem Ziel ausgewertet, Zusammenhänge aufzudecken und Ursachen von Ereignissen zu finden, um so beispielsweise Prozesse verbessern zu können. Big Data wird häufig als Oberbegriff verwendet, während Prozesse der Datenverarbeitung und -analyse als Data Mining bezeichnet werden. Statistische Datenanalysen werden schon seit Jahrzehnten beispielsweise im Qualitätsmanagement eingesetzt, um Faktoren der Prozessstabilität oder Ursachen von Qualitätsproblemen aufzuklären zu können. Was neu ist: In immer kürzerer Zeit können immer grö-

ßere Datensätze ausgewertet und kombiniert werden. Und: Auch Data Mining kann mit KI kombiniert werden. Der Name dafür: *Smart Data*. Grundsätzlich können Daten auch ohne KI statistisch analysiert werden. Und manchmal wird der Begriff der Intelligenz auch falsch eingesetzt: nämlich als Synonym für eine gigantische Rechenleistung.

2.2 Künstliche Intelligenz, maschinelles Lernen und Deep Learning – alles intelligent, aber sehr begrenzt

Was sind KI, maschinelles Lernen und Deep Learning? So werden Fähigkeiten von technischen Systemen genannt, die sich Menschen im Prinzip als Lernen und Denken vorstellen. Solche Systeme werden in der Betriebspraxis bereits häufig eingesetzt. Verglichen mit den Fähigkeiten des Menschen handelt es sich jedoch immer noch um eine sehr begrenzte Form von Intelligenz. Deshalb wird in diesen Fällen meist nicht von KI, sondern von *maschinellern Lernen* oder *Deep Learning* gesprochen.

KI, die heute in Arbeitsprozessen eingesetzt wird, hat gelernt, sehr spezifische Probleme zu lösen. Die Voraussetzung für diese Fähigkeit: Die KI kann aus Daten Muster und Zusammenhänge ‚herauslesen‘. Im Unterschied zu einer Big Data-Analyse kann KI auch unstrukturierte Daten verarbeiten und ist hierbei nicht auf Anweisungen zur Problemlösung angewiesen, die von Menschen programmiert worden sind. KI erkennt beispielsweise Bilder, Personen, Sprache oder Geräusche. Aus diesen liest KI Muster heraus, die sie in neuen Daten wiedererkennt. KI beruht zwar auf Computerprogrammen, entwickelt jedoch eigenständig Anweisungen – sie werden als Algorithmen bezeichnet –, um Fragen zu bearbeiten. Eine mögliche Frage, die KI beantworten kann: Sind das die Geräusche eines intakten Motors?

KI wird also nicht programmiert, sondern trainiert. Sie erhält eine Auswahl an Trainingsmaterial, in dem sie Muster erkennt. Im weiteren Verlauf beurteilt die KI eigenständig weiteres Trainingsmaterial und bekommt Rückmeldungen darüber, ob sie das relevante Motiv korrekt

erkannt hat. Irgendwann wird die Technik – hat sie eine ausreichende Trefferquote erreicht – ihre Aufgaben eigenständig lösen. Dann kann sie eingesetzt werden, um beispielsweise eingehende Waren zu kontrollieren, Arbeitsschritte zu beurteilen, im Warenlager das richtige Material auszuwählen oder E-Mails von Kund/-innen zu beantworten.

Wird KI menschliche Arbeit weitgehend verdrängen? Es gibt mehr gute Gründe, diese Frage zu verneinen, als sie zu bejahen. Denn KI vermag menschliche Intelligenz nicht zu kopieren, und sie ist ihr in vielerlei Hinsicht unterlegen. So ist menschliches Lernen nicht darauf begrenzt, Daten zu analysieren und Muster zu erkennen. Und: Wenn Menschen Muster erkennen, dann ist das etwas ganz anderes als bei Maschinen. Ein Beispiel: Wird KI beispielsweise mit Fotos von Last- und Personenwagen trainiert, kann die Maschine, wenn sie später unbekannte Fotos analysiert, Fehler machen, die einem Menschen nicht unterlaufen. Denn: Sind zufällig auf den Lastwagen-Fotos dunkle Regenwolken zu sehen und auf den Personenwagen-Fotos die Sonne, dann könnte dies dazu führen, dass die Maschine ein Fahrrad unter Regenwolken als Lastwagen identifiziert.

Deshalb ist zu erwarten, dass KI fähig sein wird, Teile menschlicher Arbeit zu automatisieren. Sie wird aber nie in der Lage sein, den Menschen als Entscheider, Gestalter, Optimierer und Kontrolleur zu verdrängen. Menschen haben Intuitionen, Motive, Erfahrungs- und Kontextwissen, und sie sind in der Lage, Sinn zu verstehen. Menschliches Denken ist nicht auf mathematische Operationen beschränkt. Und Menschen lernen im Unterschied zur Technik, weil sie etwas lernen wollen und weil Lerninhalte in einem sinnvollen Zusammenhang mit ihren Erfahrungen und Bedürfnissen stehen. KI ist dagegen nur dann leistungsfähig, wenn es um darum geht, Faktenwissen aufzubereiten oder aus Daten Muster herauszulesen. Doch den größeren Zusammenhang versteht KI nicht, denn ihr fehlen der „gesunde Menschverstand und Hintergrundwissen“ (Lenzen 2018: 47). Deshalb kann KI selbstverständlich erhebliche Fehlurteile produzieren: „Ein Wetterprognosesystem wird ohne zu zögern für den nächsten Tag eine Höchsttemperatur von 500 Grad ankündigen, wenn es diese Zahl in der entsprechen-

den Spalte seiner Routine vorfindet.“ (Ebd.: 47) KI wird immer zwingend auf die Zusammenarbeit mit Menschen angewiesen sein; umgekehrt gilt das nicht. Und sie wird immer versagen, wenn es um Intuition, Kreativität, das Setzen von eigenständigen Zielen und um Probleme geht, die nicht mathematisierbar sind.

2.3 Autonome Systeme – Intelligenz oder nur Automation?

Autonome Systeme sind kombinierbare technische Einrichtungen, die Arbeiten automatisieren können. Ein Teil der Autonomen Systeme wird in direkter Interaktion mit Menschen eingesetzt. Der Name dafür: *Assistenzsystem*. Der Begriff der Assistenz soll signalisieren, dass die Systeme den Menschen unterstützen. Tatsächlich ist das Verhältnis nicht so eindeutig, sondern vielschichtiger. Denn Menschen können von einem Assistenzsystem auch angeleitet, überwacht und kontrolliert werden.

Autonome Systeme bestehen aus folgenden Komponenten: Eingabe- und Ausgabegeräte, Sensorik, Aktorik, Rechner, weiterer Hardware und Software sowie der Verbindung mit der Firmen-IT. Die eingesetzte Sensorik ist vielfältig: Kamerasysteme, GPS-Empfänger oder RFID-Chips gehören dazu. Auch Smart Data und KI können, aber müssen nicht Bestandteil Autonomer Systeme sein.

In der Debatte um die Industrie 4.0 wird häufig ein Vorgang, der automatisiert abläuft, falsch als intelligent bezeichnet. Diese bewusste oder unbewusste Verwechslung hat weitreichende Folgen: Sie erklärt, weshalb heute selbst Forscher/-innen und weitere Expert/-innen davon ausgehen, dass in der Produktion überwiegend intelligente Technik eingesetzt werde und damit die Vision einer Industrie 4.0 bereits weitgehend realisiert sei. Dies ist jedoch keineswegs der Fall. Denn viele der immer wieder präsentierten Beispiele für den Einsatz einer neuen intelligenten Technik entpuppen sich bei genauer Betrachtung als schlichte Automatisierung.

Autonome Systeme können kombiniert werden, und sie sind in hohem Maße fähig, sich zu vernetzen. Damit können sie vielfältig eingesetzt werden, vor allem in Verbindung mit KI. Beispiele: Sie können Flugzeuge fliegen, Werker/-innen in Montageprozessen anleiten, Werkzeuge bei Qualitätsmängeln automatisch nachjustieren, den Wareneingang kontrollieren, eine Prozessstörung voraussagen, Anwälte/-innen bei der Lösung einer juristischen Frage und Ärzte/-innen bei der Diagnose von Röntgenbildern unterstützen.

Unternehmen setzen Assistenzsysteme ein, um Arbeitsprozesse transparent, schneller, billiger und qualitativ besser zu machen und um sie zu planen, zu steuern und zu kontrollieren. Assistenzsysteme unterstützen die unternehmensinterne Kommunikation und Koordination. Ein weiteres Ziel: die Arbeitsplätze ergonomisch zu verbessern. Die entscheidenden Unternehmensziele sind jedoch unverändert: Kosten senken, Produktivität steigern.

3. HYBRIDE ARBEITSSYSTEME – WER HAT WEM WAS ZU SAGEN?

Arbeitet der Mensch mit *konventioneller Technik*, dann ist er in der Regel derjenige, der über die Technik bestimmt und sie für Zwecke nutzt, die er festlegt. Nun gilt diese Aussage im Arbeitsprozess zwar immer nur eingeschränkt, weil das Management die übergeordneten Ziele definiert, und sich Produktionsmittel immer nur im vom Management vorgezeichneten Rahmen nutzen lassen. Doch innerhalb dieses Rahmens entscheiden die Arbeitskräfte, wie sie die Ziele erreichen wollen. Und sie verfügen zudem über eine gewisse Bandbreite an Möglichkeiten, wie sie die Technik einsetzen und steuern. Konventionelle Technik ist damit vor allem ein *Werkzeug*, das von Arbeitskräften strategisch eingesetzt wird. So wird ein Auftrag, den eine Werkzeugmaschine bearbeitet, vorab von Menschen erzeugt, in Maschinendaten übersetzt, die Maschine wird programmiert, die Bearbeitung überwacht und gegebenenfalls korrigiert. Menschen können die Systemzustände

der Maschine verstehen, und sie können mit einiger Sicherheit voraussehen, in welchem Zustand Maschine und Produkt sich in nächster Zukunft befinden werden. Konventionelle Technik ist aus Sicht des Beschäftigten: ein Objekt und kein Akteur.

Mit Autonomen Systemen ändert sich das Verhältnis von menschlicher Arbeitskraft und Technik wesentlich – der Mensch droht an Eigenständigkeit zu verlieren. Denn Technik plant und entscheidet selbst und ist autonom fähig, sich neuen Lagen anzupassen. Damit macht sie eine letztlich ungeheure Karriere vom Werkzeug des Menschen zum eigenständigen Akteur. Auf diese Weise entstehen „hybride Arbeitssysteme“, in denen menschliche und nicht-menschliche Akteure bei geteilter Verantwortung interagieren (Weyer 2006: 8). Je vielseitiger nicht-menschliche Agenten handeln und je mehr Systemzustände sie einnehmen können, je weniger die Menschen deren Aktionen verstehen, erwarten und voraussagen können, desto stärker engt die Technik den Spielraum des Menschen für sein eigenes strategisches Handeln ein. Wer richtet sich nach wem? Das wird zur entscheidenden Frage.

„In dem Maße, in dem sich die intelligente Technik der strategischen Manipulation entzieht, und damit gerade nicht mehr als Mitspieler zur Verfügung steht, verändert sich jedoch der Status des Menschen, der in zunehmendem Maße gezwungen wird, sein Verhalten, dem der Maschine anzupassen.“ (Ebd.: 20)

Konkret: Der Einsatz autonomer Technik kann damit einhergehen, dass Beschäftigte die Kontrolle über ihre Arbeitsleistung in Teilen oder in Gänze verlieren.

Aber: Technik kommt nicht einfach über die Menschen wie ein Gewitter, und sie verändert Arbeit nicht in einer Art und Weise, die ohne Alternative ist. Es sind vielmehr Menschen, welche die Technik nach ihren Kriterien und Interessen beschaffen, ihren Einsatz planen und sie in einer Arbeitsorganisation einsetzen, die wiederum von Menschen gestaltet worden ist. Technik determiniert deshalb nicht die Arbeit. Sie ist vielmehr ein Element von vielen, aus denen von Menschen das sozio-technische System gestaltet wird; sie ist also Bestandteil so-

zio-technischer Systemgestaltung. Dies bedeutet: Wer Arbeitssysteme gestaltet, gestaltet neben der Technik immer auch die Arbeitsorganisation und die Aufgaben und Funktionen des Personals. Planer/-innen gestalten die Arbeitsteilung, die Koordination von Arbeit, die Aufteilung der Funktionen zwischen Mensch und Maschine und den Arbeitsablauf. Bei diesen Gestaltungsarbeiten folgen die Planer/-innen des Betriebes immer den Leitbildern des Unternehmens. Sie orientieren sich an Vorstellungen ideal gestalteter Produktionsprozesse, ideal gestalteter Arbeitsorganisationen und idealer Mitarbeiter/-innen.

Setzt ein Unternehmen Autonome Systeme ein, dann ist das ein grundlegender Eingriff in die Gestaltung des Arbeitssystems: Denn damit findet eine *Transformation von Arbeit* statt. Arbeit wird in ihrer Funktion und Struktur verändert. Nicholas Carr erläutert das so:

„Ein arbeitssparendes Gerät substituiert nämlich nicht lediglich einen Einzelbestandteil einer Tätigkeit. Es verändert die ganze Aufgabe seinem Wesen nach, die Rollen, die Einstellungen und Fertigkeiten der beteiligten Personen inbegriffen.“ (Carr 2014: 49)

Neue Technik schafft neue Optionen, um Arbeitssysteme neu zu gestalten. Wie diese Optionen genutzt werden, darüber entscheiden die Gestalter/-innen von Arbeitsprozessen. Nicht selten verändern diese Planer/-innen zunächst die Arbeits-Organisation, um den Weg für die Automatisierung einzelner Tätigkeiten zu ebnen. Beispielsweise müssen zunächst logistische Tätigkeiten von der Produktbearbeitung abgespalten werden, um sie dann im nächsten Schritt zu automatisieren.

Digitalisierte Technik kann sehr verschieden eingesetzt werden. Mögliche Entwicklungen: Arbeit kann – gemessen am Kriterium der Qualifikation – auf- oder abgewertet oder polarisiert werden. Welches Szenario wird eintreten? Wie autonom oder abhängig ist konkret in jedem Unternehmen die Technik vom Menschen? Das entscheidet der, der das Sagen über die sozio-technische Systemgestaltung hat (vgl. Itermann et al. 2016). Denkbar ist bisher vieles (vgl. Timpe/Jürgensohn/Kolrep 2002: 47f., zitiert in Weyer 2006: 5):

- Technik kann den Menschen partiell unterstützen.
- Sie führt nur etwas aus, wenn der Mensch dies vorher bestätigt.
- Sie handelt autonom, jedoch kann der Mensch ein Veto einlegen.
- Sie handelt autonom und informiert den Menschen nur, wenn der bewusst danach fragt.
- Sie ignoriert den Menschen.

Wie bedeutend die Autonomen Systeme für den Menschen und seine Arbeit werden, das hängt also davon ab, wie Menschen die Arbeitsteilung im Unternehmen und wie sie die Interaktion von Mensch und Technik gestalten. Das ist alles andere als eine triviale Erkenntnis. Denn über Jahrzehnte prognostizierte die Arbeitsforschung, auf Dauer werde alternativlos die Fremdbestimmung im Arbeitsprozess zunehmen. Die von einer Mehrheit getragene Vorstellung war: Der Mensch wird im Prozess der Technisierung zunehmend dequalifiziert, beherrscht und kontrolliert. Im Interesse der maximalen Kapitalverwertung geht es den Unternehmen darum, menschliche Arbeit so in den Arbeitsprozess zu integrieren, dass die Arbeitsleistung unabhängig von Willen und Kooperationsbereitschaft der Arbeitskräfte abgerufen werden kann. Viele dieser analytischen Prognosen gehen auf eine Kritik der kapitalistischen Produktionsweise zurück, die in den 1970er Jahren von Harry Braverman (1974) mit Blick auf den Taylorismus zugespitzt wurde. Die Geschichte der Arbeitsgestaltung ist demnach eine Geschichte der Perfektionierung von Kontrolle. Hierbei hatte Braverman Varianten der persönlichen Überwachung und Beherrschung aufgrund inhaltlich und zeitlich streng strukturierter Arbeitsabfolgen im Blick. Spätere Ansätze betonten stärker die Rolle der Technik, von deren Auswahl und Gestaltung. Deren Annahme lautete: Unternehmen wählen die Technik, mit der sie die Beschäftigten am besten beherrschen und kontrollieren können.

Der Glaube an den Technikeinsatz als Mittel einer sich perfektionierenden Kontrolle führte im Denken zu einem *Dualismus*: Es stehen sich Verantwortung und Autonomie des Menschen und die Technik gegenüber, die dem Menschen die Autonomie entzieht. Noch in den

1990er Jahren wurde intensiv über diese Alternative diskutiert: Ist die Rationalisierung ‚anthropozentrisch‘ oder ‚technikzentriert‘? Die Forderung dieser kapitalismuskritischen Kreise: Der Mensch soll im Mittelpunkt stehen, nicht die Technik. Diese Alternative erwies sich jedoch als falsch (vgl. Böhle 1998). Es zeigt sich vielmehr, dass Technik für unterschiedliche Formen der Arbeitsgestaltung eingesetzt werden kann: nicht nur für dequalifizierte und hochgradig fremdbestimmte.

4. WELCHE CHANCEN UND RISIKEN BRINGT „INTELLIGENTE TECHNIK“ MIT SICH?

Wie die Arbeit künftig gestaltet wird, das lässt sich nicht prognostizieren. Es können bestenfalls Gestaltungskorridore skizziert werden und Annahmen über mögliche Entwicklungen. Denn die Einflüsse, die berücksichtigt werden müssen, sind ausgesprochen vielfältig (vgl. Gerst 2018). Zu ihnen gehören: das Interesse an Kapitalverwertung, Leitbilder der Arbeitsgestaltung, der jeweilige politische und rechtliche Rahmen, die Verfügbarkeit von Arbeitskräften, die Möglichkeit, den Standort ins kostengünstige(re) Ausland zu verlagern, und die Macht der Gewerkschaften.

4.1 Wie viele Arbeitsplätze gehen verloren?

Die Furcht ist weit verbreitet, dass automatisierte Technik menschliche Arbeit ersetzt. Aktuelle empirische Studien belegen hingegen für die Vergangenheit, dass es bisher kaum zu *technologisch bedingter Arbeitslosigkeit* gekommen ist. Für die Zukunft können jedoch nur Szenarien mit bestimmten Annahmen formuliert werden, die sich als falsch herausstellen können. Wer die Annahmen der Studien nicht mitbedenkt, kann leicht in die Irre gehen.

So wird die weltweit publizierte Studie von Carl B. Frey und Michael A. Osborne (2013) fast immer falsch interpretiert. Sie untersucht, inwieweit Roboter und Computer menschliche Arbeit ersetzen können.

Die Autoren lassen jedoch außer Acht, dass der Automatisierung auch immer Grenzen gesetzt sind. Das heißt, nicht alle Arbeits- und Produktionsprozesse, die technisch und organisatorisch automatisiert werden könnten, werden tatsächlich auch automatisiert. Dieser wesentliche Punkt wird von den meisten Leser/-innen der Studie übersehen. So wird diese Studie noch heute regelmäßig als Beleg angesehen, dass das Risiko der Automatisierung ganzer Berufe hoch sei. Übersehen wird, dass für Frey und Osborne mehrere wichtige Fragen gar keine Rolle spielen: wie teuer die Automatisierung für die jeweiligen Unternehmen werden wird, ob also die Automatisierung die menschliche Arbeit tatsächlich mit Kostenvorteilen oder wenigstens kostendeckend ersetzen kann; ob Kund/-innen diese Umstellung von Mensch auf Maschine akzeptieren; ob sich tatsächlich Berufe in Gänze ersetzen lassen oder nur konkrete Tätigkeiten des jeweiligen Berufes; ob wegen der Investitionen in automatisierte Technik oder wegen der so verursachten Produktivitätsgewinne neue Arbeitsplätze entstehen. Da die Autoren das alles nicht berücksichtigen, wundert es nicht, dass sie zu dem Ergebnis kommen, bereits heute seien 47 Prozent der Beschäftigten in den USA von Computer und Roboter bedroht.

Deshalb kommen auch die Studien zu ganz anderen Ergebnissen, welche die Grenzen der Automatisierung ebenso wie Potenziale für neue Beschäftigung berücksichtigen. Nach einer Untersuchung des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) (vgl. Dauth et al. 2017) hat ein Roboter in der Vergangenheit durchschnittlich zwei Arbeitskräfte in der Industrie ersetzt, während im gleichen Zeitraum etwas mehr als zwei Arbeitsplätze außerhalb der Industrie geschaffen wurden. Der Beschäftigungsverlust wird in Gänze also überkompensiert. Dieser Ausgleich geht jedoch mit einem Wandel der Beschäftigten- und Wirtschaftsstruktur einher.

Gegen die Studie von Wolfgang Dauth et al. (ebd.) ließe sich einwenden, sie untersuche nur den Einsatz von Robotern, nicht jedoch den der KI, die eventuell noch gewichtiger bewertet werden müsse. Denn KI könne zudem im Dienstleistungsbereich eingesetzt werden, und sie sei zudem in der Lage, Aufgaben zu automatisieren, die zumindest in

der Vergangenheit von Robotern nicht automatisiert werden konnten. Beispiele: alle Tätigkeiten in den Bereichen der Administration, der Planung und Steuerung. So sind auf KI basierende Maschinen in der Lage, E-Mails zu formulieren und zu versenden, Rechnungen zu bearbeiten und Gespräche mit Kund/-innen zu führen. Diese Beispiele zeigen: Es bedarf Studien, die von einem breiteren Verständnis digitaler Technik ausgehen. Dieses findet sich in einer Untersuchung des Zentrums für Europäische Wirtschaftsforschung (ZEW) (vgl. Arntz/Gregory/Zierahn 2018), die sich mit den Folgen von Investitionen in moderne digitale Technik befasst. Das Ergebnis: Die neue Technik sei fähig, manuelle und vor allem geistige Routinearbeit zu ersetzen. In einem geringeren Ausmaß könne sogar manuelle Arbeit verdrängt werden, die nicht standardisiert sei. Analytische und interaktive Arbeit nehme dagegen zu. Vorausgesetzt die Wirtschaft wachse, führe diese Entwicklung bislang in Gänze zu keinen Arbeitsplatzverlusten.

Laut einer weiteren Studie des IAB (vgl. Zika et al. 2018) werden durch die Digitalisierung von 2020 bis 2035 einerseits 1,46 Millionen Arbeitsplätze ersetzt und andererseits 1,4 Millionen neue Arbeitsplätze geschaffen werden. Der Verlust an Arbeitsplätzen fällt also mit 60.000 Stellen moderat aus, die Struktur der Beschäftigung wird sich jedoch stark wandeln. Die größten Verluste wird es in Berufen der Administration und der kaufmännischen Dienstleistungen (minus 268.000) geben, und in denen, die Maschinen und Anlagen steuern und warten (minus 278.000). Die größten Zuwächse erwarten die Autoren in den Bereichen der Medien und der geistes- und sozialwissenschaftlichen und künstlerischen Berufe (plus 161.000) und der Berufe in IT und Naturwissenschaften (plus 149.000). Regional werden die Folgen entsprechend der Branchenstruktur variieren.

Damit stimmen die aktuellen Studien in drei bedeutenden Punkten überein: Die Beschäftigungsverluste werden in Gänze gering oder sogar sehr gering sein. Und: Die Arbeitsmärkte werden sich sehr dynamisch wandeln. Zudem gibt es Hinweise auf eine sich verschärfende Polarisierung: Vor allem hochqualifizierte Beschäftigte werden von dieser Modernisierung profitieren. Zum einen wird es mehr solcher

Arbeitsplätze geben, und deren Entgelte werden sich erhöhen. Dagegen wird sich wegen des erhöhten Einsatzes digitalisierter Technik das Entgelt von gering Qualifizierten nicht verändern, das von Beschäftigten auf dem Qualifikationsniveau eines Facharbeiters/einer Facharbeiterin jedoch verringern.

4.2 Kommt die Technik ohne den Menschen aus? Über die Entwicklung der Anforderungen an Qualifikation und Kompetenz

Mit Autonomen Systemen treiben Unternehmen die Verwissenschaftlichung der Produktion voran. Hierbei wird Wissen mit der Folge in die Technik integriert, dass die Beschäftigten dieses Wissen weniger oder gar nicht mehr benötigen und bereithalten müssen. Digitalisierung ist deshalb mit einem Kompetenzverlust verbunden. Werden Roboter und automatisierte Werkzeugmaschinen eingesetzt, dann verlieren Arbeitskräfte Wissen darüber, wie die Produkte bearbeitet werden und auf welche Feinheiten es bei der Bearbeitung ankommt. Sie verlieren so auf Dauer an Intuition und Erfahrung, können sich den Arbeitsprozess also subjektiv immer weniger sinnlich aneignen (vgl. Böhle et al. 2011): Sie verlieren buchstäblich an Gefühl für die Arbeitsschritte, und ihr Wissen wird geringer, wie beispielsweise Werkstücke und Arbeitsmittel reagieren, wenn der Mensch in den Produktionsprozess eingreift. Dieses subjektive Arbeitshandeln ist auch im Dienstleistungssektor unverzichtbar, etwa im direkten Kontakt mit Kund/-innen. Auch dieses Wissen und die damit verbundene soziale Kompetenz schwinden, wenn die Kommunikation weitgehend über digitale Assistenzsysteme geführt wird.

Ob Menschen von Autonomen Systemen und der KI in der Gesamtbilanz dequalifiziert werden oder nicht, ist pauschal nicht zu beantworten. Trotzdem wird häufig die These vertreten, das Wissen der Arbeitskräfte werde weitgehend in die Technik verlagert, weshalb die Menschen zumindest Teile ihrer Kompetenz einbüßen würden. Beispiele dafür gibt es: Beschäftigte, die bei der sogenannten Werkerfüh-

rung von technischen Assistenzsystemen Schritt für Schritt angeleitet und kontrolliert werden; die Picker in der Kommissionierung; Büropersonal, das wegen der Digitalisierung nur noch einen Bruchteil eines einzelnen Kundenauftrags bearbeiten darf. Die Digitalisierung zerstückelt also viele Tätigkeiten und entleert sie von Kompetenzen – ein Prozess, der als „Amazonisierung von Arbeit“ (Butollo/Ehrlich/Engel 2017: 1) bezeichnet werden kann.

Es gibt keine Pauschallösung, wie Funktionen zwischen Technik und menschlicher Arbeit aufgeteilt werden können. Es hängt vor allem von den Merkmalen der Arbeitstätigkeit ab. Einfachste Tätigkeiten lassen sich heute von Maschinen steuern und überwachen. Dort, wo Technik komplexe Prozesse steuert, ist zwar einerseits sehr viel an menschlichem Wissen in die Technik implementiert, andererseits gewinnt menschliche Arbeit aber auf dieser höheren Fertigungsstufe an systemrelevantem Produktionswissen (vgl. Malsch 1987). Das Produktionspersonal benötigt zwar keine detaillierten Kenntnisse des Produktionsprozesses mehr, es muss aber lernen, die automatisierte Bearbeitung zu überwachen, zu steuern und korrigierend einzugreifen. Korrigieren musste der Mensch auch früher schon, als Automationstechnik eingesetzt worden ist. Trotzdem gab es bereits in der Vergangenheit den weit verbreiteten Irrtum, Technik komme ohne den Menschen aus. Die Systeme waren aber immer schon nur scheinbar autonom. Scheinbar, weil die korrigierenden Eingriffe von Menschen für Außenstehende überwiegend nicht sichtbar waren und sind. Soll die Produktion reibungslos funktionieren, kommt es jedoch auf diese Eingriffe an. Und sie sichern der menschlichen Arbeit Kompetenz und Autonomie (vgl. Böhle 1998).

Heute wird manchmal bezweifelt, ob die immer leistungsfähigeren Autonomen Systeme dieser menschlichen Korrekturarbeit überhaupt noch bedürfen. Studien zeigen jedoch, dass Autonome Systeme gar nicht in der Lage sind, Arbeitsprozesse dauerhaft und lückenlos unabhängig von menschlicher Kompetenz auszuführen. Die Debatte über die „Ironie der Automatisierung“ (Bainbridge 1983: 1) illustriert das. Von Ironie wird deshalb gesprochen, weil es Folge der Automatisie-

rung einfacher Tätigkeiten ist, dass schwierige Tätigkeiten noch schwieriger werden, was die Sicherheit von Prozessen gefährdet. Das Argument: Versagt die Technik, dann kommt es doch wieder auf den Menschen an. Der ist jedoch im Störfall unter Umständen überfordert, weil er wegen der Automatisierung an Kompetenz verloren hat. Die Beschäftigten sind vor allem dann überfordert, wenn der Arbeitseinsatz so geplant wird, dass ihre Kompetenzen nur im Störfall gefordert sind, sie im Normalfall aber gar nichts tun müssen. Die Folgen: Beschäftigte müssen mit gezielten Trainings auf Störfälle vorbereitet werden. Und die Produktion muss so gestaltet werden, dass Menschen sinnvoll immer wieder in eigentlich durchautomatisierte Prozesse eingreifen müssen.

Die Konsequenz aus diesen Erfahrungen und Überlegungen: Menschen müssen weiterhin fähig sein, Prozesse zu steuern und bei Technikversagen korrigierend einzugreifen. Und dies erfordert spezifische neue Kompetenzen. Arbeitskräfte müssen lernen, Rückmeldungen der Technik auf ihre Plausibilität hin kritisch zu beurteilen. Und sie müssen wissen, was ihnen die Technik über die diversen Bildschirme und Anzeigen über die Arbeitsprozesse nicht zu berichten weiß. Ähnliches gilt auch für Dienstleistungen. Menschen werden nicht angemessen mit Kund/-innen und Geschäftspartner/-innen umgehen können, wenn die Kommunikation nur noch vermittelt über den Bildschirm geführt wird, orientiert an Standardabläufen und elektronischen Formularen. Arbeitskräfte, die im Umfeld komplexer Prozesse tätig sind, werden deshalb in Zukunft eher höher qualifiziert sein müssen.

Welche Kompetenzen werden Beschäftigte künftig haben und haben müssen? In der Debatte über diese Frage wird häufig an alte Kontrolltheorien angeknüpft. Danach haben Unternehmen ein Interesse daran, Arbeitsprozesse so zu gestalten, dass die Beschäftigten die geringsten Qualifikationen benötigen. Zahlreiche Beispiele zeigen jedoch, dass die Unternehmen ihre Technik nicht nach dem Kriterium der Dequalifizierung auswählen; sie ist offenkundig kein unternehmerisches Ziel. Unternehmen wählen vielmehr die Technik, die ihre produktionstechnischen Ziele am besten unterstützt. Und diese Ziele sind,

zumindest in Deutschland: möglichst sicher, kostengünstig und flexibel eine möglichst variantenreiche Produktpalette herstellen. Jedoch sind diese Ziele mit einer dequalifizierten Belegschaft heute immer weniger zu erreichen. Diese entscheidende Voraussetzung wird jedoch von der Mehrheit der Unternehmens-Planer/-innen und Prozessgestalter/-innen offenkundig nicht wahr- und ernstgenommen. Ein Grund: Die planenden Ingenieur/-innen sind keine Expert/-innen der sozio-technischen Systemgestaltung. Sie handeln als Ingenieur/-innen, deren Interesse darin liegt, den Produktionsprozess aus der Perspektive der Technik gut zu gestalten.

4.3 Wer kontrolliert den Arbeitsprozess? – die Macht der Algorithmen

Die stärkste Furcht der Menschen: Wegen der KI verlieren wir Autonomie und Macht. Arbeitsforscher/-innen sind bisher mehrheitlich davon ausgegangen, dass Kontrollverhältnisse immer noch Spielräume enthalten, die genutzt werden können, um kleine Freiheiten auszuhandeln. Der Grund: Auch strikt bürokratisch geplante Prozesse funktionieren nie reibungslos. Denn es ist unmöglich, sämtliche Rahmenbedingungen zu berücksichtigen und gar deren Veränderungen vorherzusehen. Auch das Kontrollpersonal ist zu wenig mit den Details der Arbeitsprozesse vertraut, um rechtzeitig und effizient auf Abweichungen reagieren zu können. Aus diesen Gründen gibt es in Betrieben eine ‚doppelte Wirklichkeit‘: die Welt der bürokratischen Planung und die Welt der von den Beschäftigten unsichtbar korrigierten Planung (vgl. Wetz 1999).

In diesem Sinne hatte schon Cornelius Castoriadis (1997) kritisch auf die These von Karl Marx geschaut, der davon ausging, der Mensch werde im kapitalistischen Arbeitsprozess zunehmend fremdbestimmt. Er setzte dieser These entgegen:

„So sehr die Verdinglichung eine wesentliche Tendenz des Kapitalismus ist, kann sie sich in der Wirklichkeit doch niemals vollständig durchsetzen. Könnte

sie es – gelänge es dem System also tatsächlich, die Menschen in Dinge zu verwandeln, die einzig und allein von ökonomischen Kräften bewegt werden –, so würde es nicht erst auf lange Sicht, sondern augenblicklich zusammenbrechen.“ (Ebd.: 31)

Und:

„Der Kapitalismus kann nur funktionieren, wenn er ständig die im eigentlichen Sinne *menschliche* Tätigkeit der ihm unterworfenen Subjekte in Anspruch nimmt, während er zur gleichen Zeit versucht, diese Tätigkeit auf ein Mindestmaß herunterzusetzen und zu entmenslichen.“ (Ebd.: 31., Herv. i.O.)

Heute stellt sich die Frage, ob mit der KI das Kontrollverhältnis zu Lasten der Beschäftigten verändert werden kann. Als Beispiel wird häufig auf das Personal in der Kommissionierung verwiesen, das wegen des Einsatzes digitaler Techniken nahezu sämtliche Freiheitsgrade verliert. Auch die neuen engeren Grenzen, die dem Menschen im Dienstleistungssektor gesetzt werden, werden als Beispiel angeführt. Wenn die Kommunikation mit Kund/-innen nur noch digital vermittelt geführt wird, erzwingt die Maschine aufgrund dieser standardisierten Prozesse vom Beschäftigten bestimmte Verhaltensweisen, die neue Widersprüche und neue Belastungen mit sich bringen. So kann sich das Personal beispielsweise nicht mehr flexibel an die Kommunikation und Wünsche der Kund/-innen anpassen (vgl. Böhle/Stöger/Wehrich 2015). Auch wenn diese Beispiele mit Recht für die These von strikteren Kontrollverhältnissen angeführt werden können, so gilt die These, Autonome Systeme binden Menschen so in Arbeitsprozesse ein, dass sie sich in Dinge verwandeln, keinesfalls für die meisten Beschäftigungsverhältnisse.

Aber: Wird sich die Fremdbestimmung der Arbeitskräfte dann erhöhen, wenn künftig Arbeitsprozesse zunehmend via Algorithmen gesteuert werden? Über diese Frage forschen zwar bereits Wissenschaftler/-innen, jedoch sind sie bisher kaum über die Formulierung von Fragen und Hypothesen hinausgekommen; diese wissenschaftlichen Vor-

arbeiten haben zudem den Nachteil, dass es ihnen an der notwendigen Distanz zu den alten Kontrolltheorien mangelt.

Ausgangspunkt dieser ‚Software Studies‘ ist die Frage: Sind Algorithmen frei von Ideologie, Kultur und Politik? Das wird in den Studien bezweifelt und stattdessen angenommen, Algorithmen seien Teil einer Beziehung von Macht und Wissen und damit Bestandteil von Herrschaftssystemen (vgl. Goffey 2008: 19). Wie diese Beziehung konkret funktioniert, ist jedoch noch nicht erforscht.

Bisher kann nur Allgemeines darüber gesagt werden, wie Algorithmen im Arbeitsprozess wirken: Sie verändern die Realität der Arbeitsbedingungen, und sie sind Teil sozialer Verhältnisse. Beschäftigte müssen deshalb das Wirken der Algorithmen in ihren Strategien berücksichtigen. Dies kann als Macht begriffen werden, die von den Algorithmen ausgeht. Es stellt sich jedoch die Frage: Wandeln die Algorithmen, mit der von ihnen ausgehenden Macht, die Kontrollverhältnisse grundlegend oder nicht? Die vorläufige Antwort: Der Algorithmus könnte Bestandteil eines Kontrollverhältnisses sein, das *unpersönlich* ist.

In der Vergangenheit gab es Spielräume innerhalb von Kontrollverhältnissen, weil die Verhältnisse Ergebnis von *persönlichen Aushandlungen* zwischen Beschäftigten und Aufsichtspersonal waren. Beschäftigten wurden Entscheidungsspielräume zugebilligt, weil Planer/-innen und Vorgesetzte darauf angewiesen waren, die Beschäftigten zu freiwilligen korrigierenden Leistungen zu motivieren. Wenn das Kontrollverhältnis nun unpersönlich wird, verschwinden zwangsläufig auch die Aushandlungsspielräume. Und selbst wenn Beschäftigte über Handlungsspielräume verfügen, könnten Algorithmen überwachen, ob Menschen die ihnen verbliebenen Spielräume im Interesse der Unternehmensziele einsetzen oder nicht. KI bietet also diese neue Option: Die Beschäftigten haben bei ihrer Arbeit zwar persönliche Spielräume, aber Algorithmen überprüfen, wie die Beschäftigten ihren Spielraum konkret nutzen (vgl. Raffetseder/Schaupp/Staab 2017). Das käme einer erheblichen *Zuspitzung einer indirekten Steuerung* gleich. Arbeit wird zwar nicht mehr im Detail überwacht, sie wird aber – in diesem Fall

mithilfe von Algorithmen –, daran gemessen, inwieweit sich die Beschäftigten an Zielen orientieren, die das Unternehmen definiert. Die Vorteile aus unternehmerischer Sicht: die Kontrollkosten sinken, die Kontrolle wird dichter. Und vor allem: Die Beschäftigten verwandeln sich – beobachtet von Algorithmen – in einen effizienten Planer und eine effiziente Kontrolleurin ihrer eigenen Leistung.

Doch auch diese erweiterte Kontrolle hat ihre Grenzen. Je autonomer Technik agiert, desto größer werden die Prozessrisiken, die von ihr ausgehen. Auf die Kreativität, Kompetenz und Handlungsautonomie der Menschen können Prozessgestalter/-innen nur verzichten, wenn Technik immer ausschließlich die richtigen Informationen verarbeitet und immer alle relevanten Kontextfaktoren berücksichtigt. Da die Entwicklung der Technik von dieser Fähigkeit weit entfernt ist, wird auch in Zukunft der souverän handelnde Mensch im Arbeitsprozess unverzichtbar sein; weitgehend fehlerfreie autonome Technik gibt es nur für Nischen innerhalb von Arbeitsprozessen.

4.4 Erleichtert KI die Arbeit? – der Wandel der Arbeitsbelastungen

Mit der digitalisierten Technik wird die Hoffnung verbunden, eintönige Arbeit lasse sich ersetzen und dem Mensch bleiben nur die höherwertigen Aufgaben. Listen ausfüllen, Lagerbestände verwalten, Dinge und Daten suchen, sortieren und ablegen – das sei bald Vergangenheit. Menschen konfigurieren nur noch Systeme, überwachen und korrigieren sie. Auch in diesem Zusammenhang wird der KI eine besondere Bedeutung zugeschrieben. Ein Beispiel für kognitive Entlastungen sind Expertensysteme, die Informationen zur Verfügung stellen, die optimal zum jeweiligen Bedarf des/der Beschäftigten passen. Ein Beispiel für körperliche Entlastung: aktive Exoskelette, die Daten über ihren Träger/ihre Trägerin sammeln und lernen, dessen/deren persönliche Bewegungsabläufe zu unterstützen.

Insbesondere im Bereich der *kognitiven Ergonomie* sind Fortschritte denkbar. Wer Fehler macht, für die er/sie verantwortlich ist, ist ge-

stresst. Es ist deshalb im Interesse des/der Beschäftigten, von einer mit KI gestützten automatisierten Qualitätskontrolle unterstützt zu werden, um Fehler, und damit auch Stress zu vermeiden. KI ist auch fähig, akustische in visuelle und visuelle in akustische Informationen umzuwandeln – ein Beitrag, um Hör- oder Sehgeschädigte in den Arbeitsprozess zu integrieren.

Das alles ist realistisch. Es ist jedoch nicht das vorrangige Ziel von Unternehmen, die Ergonomie zu verbessern, sie wollen vor allem Produktivität und Gewinne erhöhen. Investitionen in die Ergonomie sind erst dann zu erwarten, wenn das bereits bestehende Regelwerk des Arbeitsschutzes in den Betrieben konsequent angewendet wird. Nach dem Arbeitsschutzgesetz, der Betriebssicherheitsverordnung und den Technischen Regeln zur Betriebssicherheitsverordnung dürfen nur Techniken verwendet werden, welche die Gesundheit der Beschäftigten und deren Wohlbefinden nicht gefährden. Deshalb hat der Arbeitgeber/die Arbeitgeberin entsprechende arbeitswissenschaftliche Erkenntnisse und den Stand der Technik zu berücksichtigen. Diese Regelwerke sind allerdings in der Praxis nur lückenhaft bekannt. Zudem scheitert ihr Einsatz häufig, weil die Durchsetzungsfähigkeit der Betriebsräte zu gering ist.

Ich halte fest: Neue Techniken können entlasten, aber zugleich neue Belastungen mit sich bringen. So können schlecht gestaltete Benutzerschnittstellen, unübersichtliche Informationsdarstellungen oder unberechenbares Systemverhalten zu „Technikstress“ (Hoppe 2010: 53) führen. Darüber hinaus wachsen mit dem Grad der Automatisierung die Risiken von Monotonie und psychischer Ermüdung. Selbst Arbeitsmittel, die auf den ersten Blick große Hoffnungen wecken, können mit neuen Risiken verbunden sein. Ein Exoskelett entlastet die Armmuskeln – hat aber ein Eigengewicht von fünf bis zehn Kilogramm, das die Beschäftigten zu tragen haben. Das Exoskelett ist aber nicht nur schwer, es schränkt auch den Bewegungsspielraum ein – beides erhöht wiederum die Unfallgefahr. Ein weiterer möglicher Nachteil: Das Gerät könnte dem Unternehmen Anlass sein, sinnvolle Investitionen in ergonomische Fortschritte zu unterlassen. Möglicherweise

ist der Einsatz des Exoskeletts für einen Betrieb billiger als Maßnahmen, um Arbeitsschritte über Herz- und Kopfhöhe zu vermeiden.

Im Bürobereich kann KI ermüdende Routinearbeit ersetzen. Aber auch hier ist die ergonomische Bilanz nicht eindeutig. Denn mit dem Einsatz neuer digitaler Assistenten könnte der Arbeitgeber/die Arbeitgeberin erwarten, dass nun schneller, effizienter und standardisierter gearbeitet wird. Damit können erhöhte Leistungserwartungen ergonomische Vorteile sogar zunichtemachen. Ähnliches gilt für die flexiblen Arbeitsformen, welche die neuen Techniken möglich machen. So ist denkbar, dass Büroarbeit der Zukunft inhaltlich an Attraktivität gewinnt, aber zugleich mit ständiger Erreichbarkeit, Multitasking und Selbstausbeutung verbunden ist.

Für Produktion wie Büroarbeit zeichnet sich gleichermaßen ab, dass betriebliche Planer/-innen für die Zeiten der Modernisierung zu geringe personelle und zeitliche Ressourcen vorsehen. So sind für die Beschäftigten Phasen der Reorganisation meist zugleich Phasen der Überlastung. Da sich Reorganisationen häufen, kann daraus eine *dauerhafte Überlastung* entstehen.

4.5 Folgen für Überwachung und Persönlichkeitsrechte

Wo Autonome Systeme eingesetzt sind, werden immer mehr Beschäftigtendaten ermittelt und verarbeitet. Das kann von Unternehmen als Basis für noch striktere Überwachung, Verhaltenskontrolle und Personalauswahl genutzt werden. Selbst die Technik, welche die Beschäftigten unterstützt, überwacht letztlich, wenn sie so konzipiert ist, dass sie sich dem individuellen Bedarf anpassen kann; das gilt beispielsweise für Systeme, die das persönliche Lernen unterstützen.

Oft werden Systeme eingesetzt, um Produktionsabläufe automatisch zu steuern. Diese erfassen Störungszeiten, Fehler und Leistungsdaten, damit auch Daten über das Verhalten der Beschäftigten. Sogenannte Enterprise-Resource-Planning(ERP)- und Manufacturing-Execution(MES)-Systeme steuern automatisch Kapazitäten, erzeugen Aufträge, bestellen Material und analysieren Prozessdaten (vgl. Gröger

2018). Dass diese Systeme Beschäftigte unter Druck setzen, ist nicht neu. Denn auch ältere MES- und ERP-Systeme verarbeiten Rückmeldedaten – auch über die Arbeit der Beschäftigten – aus den Produktionsprozessen; die Beschäftigten werden in den Datensätzen übrigens als Ressourcen geführt. Viktor Steinberger macht darauf aufmerksam, dass mit dieser neuen Datenmenge auch eine neue Qualität erreicht wird: „Das qualitativ Neue liegt in der Zusammenführung so vieler unterschiedlicher Informationen zu einem immer dichteren und zeitlich verfügbaren Datenbild.“ Was Steinberger (2011: 18) hier für neue Systeme der Auftrags- und Kapazitätssteuerung festhält, gilt in einem gesteigertem Maß für Smart Data Analysen und den Einsatz von Assistenzsystemen. Diese Systeme bieten noch viel mehr Möglichkeiten, Daten über Beschäftigte zu erfassen, zu speichern und zu verarbeiten. Unternehmen können solche Auswertungen nutzen, um beispielsweise ihr Personal nach Leistungsfähigkeit auszusortieren. Personalverantwortliche, die nur noch die Arbeitnehmer/-innen mit den wenigsten Fehlern und kürzesten Pausen beschäftigen wollen, erhalten so die Daten, die sie dafür brauchen.

Das deutsche Recht zum Beschäftigtendatenschutz setzt jedoch Grenzen: Es schließt vieles aus, was an Überwachung und Datenverarbeitung bereits möglich ist (vgl. Martini/Botta 2018), vor allem die Erstellung von detaillierten Persönlichkeitsprofilen. Wer Daten der Beschäftigten erfasst und verarbeitet, darf dies nur für Zwecke tun, die vorab genau beschrieben werden. Konkret: Die Datenverarbeitung muss für diese Zwecke erforderlich und sie muss zudem im beabsichtigten Umfang angemessen sein. Um die Angemessenheit zu überprüfen, ist der Arbeitgeber/die Arbeitgeberin verpflichtet, die beabsichtigte Datenverarbeitung gegenüber den Grundrechten der betroffenen Personen abzuwägen. Zu den Grundrechten zählen in diesem Zusammenhang das Recht auf *informationelle Selbstbestimmung* und der *Schutz der menschlichen Würde*. Das Datenschutzrecht fordert, Missbräuchen vorzubeugen, etwa indem der Technik gar nicht erst erlaubt wird, bestimmte Daten zu erfassen, indem Daten anonymisiert und nach Zweckerfüllung gelöscht werden.

Sind Überwachungsmöglichkeiten jedoch erst einmal vorhanden, ist das Risiko gegeben, dass sie auch genutzt werden. Das befürchten Mario Martini und Jonas Botta am Beispiel von Exoskeletten, die sich an den Träger anpassen:

„Indem ein aktives Exoskelett seinen Träger bei der Erfüllung seiner betrieblichen Aufgaben gleichsam wie ein Schatten folgt, kann es insbesondere unbemerkt die Tür zu einem in seiner Persönlichkeit ausgeleuchteten Beschäftigten am panoptisch überwachten Arbeitsplatz aufstoßen.“ (Ebd.: 636)

Werden Möglichkeiten zur Überwachung missbraucht, sind die Konsequenzen gravierend: Persönlichkeitsrechte werden verletzt, und die Beschäftigten sind einem wachsenden Überwachungs- und Anpassungsdruck ausgesetzt. Entscheiden technische Systeme auf der Grundlage von Beschäftigtendaten automatisiert über personelle Maßnahmen, kann eine *Diskriminierung* vorliegen. Die Risiken, die mit diesen Techniken verbunden sind, sind also beachtlich. Es ist entscheidend, ob und wie der Beschäftigtendatenschutz diesen Herausforderungen gerecht werden kann. Zumal wir davon ausgehen können, dass Unternehmen die Gewinne, die Smart Data verspricht, mehr am Herzen liegen als der Schutz der Daten der Beschäftigten.

5. MENSCH UND MASCHINE ALS TEAM – WIE KANN DIE ARBEIT MIT AUTONOMEN SYSTEMEN GESTALTET WERDEN?

Mit Autonomen Systemen und KI wird vielfältig gearbeitet. Deshalb gibt es für die Frage, wie die Arbeitssysteme gestaltet werden sollen, keine pauschalen Lösungen. Wir verfügen heute über einen bereits großen Fundus an Gestaltungswissen. Studien belegen, dass die Arbeitssysteme leistungsfähiger sind, bei denen Mensch und Technik als Team zusammenarbeiten. Wie Arbeitssysteme tatsächlich gestaltet werden, das ist aber auch eine Machtfrage.

5.1 Gestaltungswissen und Gestaltungsmacht – mehr Autonomie oder mehr Kontrolle?

Der sozio-technische Systemansatz (vgl. Grote et al. 1999) liefert das Handwerkszeug, mit dem Arbeitssysteme zugleich leistungsfähig und human gestaltet werden können. Es gibt auch fundierte arbeitswissenschaftliche Leitbilder, die genutzt werden können, um digitalisierte Arbeit menschengerecht zu gestalten. Also Grund genug für *Gestaltungsoptimismus*? Hinzu kommt, dass der Technikdeterminismus heute als überwunden gilt: Niemand vertritt mehr die These, Technik erzwingt bestimmte Formen von Arbeit. Vielmehr wird die Debatte von der These geprägt, die Folgen der Technik für die Arbeit seien gestaltbar. Was jedoch ungewiss ist: Wie groß sind die Spielräume, um die Ökonomie so zu gestalten, dass sie dem Menschen dient?

Ein naiver Gestaltungsoptimismus könnte von der Devise ausgehen, es brauche ein Leitbild und guten Willen, dann wende sich alles zum Besseren. So einfach ist die Praxis aber nicht zu verändern. Denn Unternehmen orientieren sich vorwiegend am Ziel einer kurzfristigen Gewinnmaximierung. Darüber hinaus ist betriebliche Arbeitspolitik nicht notwendigerweise rational. Unternehmen können sich in der Einschätzung irren, wie sie die Arbeitskräfte einzusetzen haben, um ihre Gewinne zu optimieren. Betriebswirtschaftliche Lehrbücher raten dazu, bei komplexen und kapitalintensiven Arbeitsprozessen den Anteil qualifizierter und souverän handlungsfähiger Arbeitskräfte eher zu erhöhen – wie viele Unternehmen diesen Rat befolgen, ist längst nicht ausgemacht. Zumal heute in der Praxis das Bestreben überwiegt, Arbeit zunehmend zu verwissenschaftlichen und von nicht kontrollierbaren subjektiven Beiträgen der Beschäftigten zu säubern (vgl. Böhle et al. 2011: 23).

Aber: Auf Dauer wird eine Politik der zunehmenden Kontrolle und Dequalifizierung scheitern. Die grenzenlose Kontrolle der Beschäftigten hat sich bereits in der Geschichte der Arbeitsgestaltung als Illusion erwiesen. Warum soll das in der Zukunft anders sein? Ich gehe deshalb davon aus, dass die arbeitspolitischen Trends – auf Grundlage der

Überlegungen von Castoriadis – künftig zwischen zwei sich widersprechenden Zielen pendeln werden: zwischen der Intensivierung von Kontrolle und der Autonomisierung menschlicher Arbeit.

Bei der Entscheidung darüber wird – neben Moden und Leitbildern – vor allem die Verteilung der Macht den Ausschlag geben. Die Konflikte werden auf verschiedenen Ebenen ausgetragen, vor allem im Betrieb, wo die Arena der Auseinandersetzung von der Politik festgelegt ist und wird; dieser Rahmen umfasst bisher das Arbeitsrecht, den Datenschutz, den Gesundheitsschutz und die Mitbestimmung. Deshalb ist dieser Rahmen auch umstritten. So richten Gewerkschaften wie Arbeitgeberverbände weitere Forderungen an den Staat. Aktuell werden diskutiert: der Stellenwert von Prävention im Gesundheitsschutz; der Grundsatz der Missbrauchsvermeidung im Beschäftigtendatenschutz; eine Mitbestimmung, bei der Arbeitgeber/-innen tatsächlich mit Betriebsräten Vereinbarungen treffen müssen. Debattiert wird auch über eine weitere Flexibilisierung des Arbeitszeitgesetzes; in der Hauptsache geht es darum, inwieweit Flexibilität im Interesse der Beschäftigten oder der Betriebe genutzt werden kann. Arbeitgebervertreter/-innen drängen in all diesen Punkten zu Rückschritten, ebenso ein Arbeitskreis von Personalvorständen großer deutscher Unternehmen (vgl. Jacobs/Kagermann/Spath 2017). Neu ist der Versuch, die Mitbestimmung nach dem Betriebsverfassungsgesetz zu schwächen und dies als Fortschritt in Richtung eines ‚demokratischen Unternehmens‘ umzu-deuten.

Wie die Arbeit gestaltet wird, das hängt letztlich davon ab, wie mächtig und durchsetzungsfähig die Betriebsräte in den jeweiligen Unternehmen und die Gewerkschaften in den jeweiligen Branchen sind. Die Handlungsmacht der Betriebsräte hängt von rechtlichen Grundlagen, der Solidarität der Beschäftigten, vom Organisationsgrad der Gewerkschaften und vom Zusammenwirken von Betriebsräten und Gewerkschaften ab. Wo es keine starken Betriebsräte und keine Tarifbindung gibt, sind die Beschäftigten auf den guten Willen und auf die Rationalität der Unternehmen im Umgang mit ihnen angewiesen. Be-

schäftigten in Engpassberufen mag es gelingen, ihre Marktmacht zu nutzen, anderen wird es an Durchsetzungsmacht fehlen.

Betriebsräte sind potenziell einflussreiche Akteure, wenn es darum geht, die Arbeit mit Autonomen Systemen zu gestalten. Sie verfügen mit dem Betriebsverfassungsgesetz, dem Arbeitsschutzrecht und dem Beschäftigtendatenschutz über weitreichende Möglichkeiten der Mitgestaltung und Mitbestimmung. Diese Rechte in der betrieblichen Praxis zu nutzen und durchzusetzen, ist aber alles andere als einfach und selbstverständlich.

So lautet der wohlbegründete Rat an die betriebliche Interessenvertretung: *frühzeitig mitgestalten*. Doch fehlen Betriebsräten oft die notwendigen Informationen über Modernisierungsvorhaben, Manager/-innen sind nicht immer auskunftsfreudig und gesprächsbereit. Dies muss keine strategische Absicht sein, die sich bewusst gegen Betriebsräte richtet. Denn viele Unternehmen pflegen das Prinzip eines dezentralen Managements mit eigener Budgetverantwortung. Die Folge davon: eine Vielfalt an unverbundenen Reorganisationsprojekten. So fehlt es oft nicht nur aus Sicht der Betriebsräte, sondern auch aus der des Managements an Transparenz und zentraler Steuerung. Betriebsräten fehlen deshalb oft auch kompetente und entscheidungsfähige Ansprechpartner/-innen im Management. Interessenvertreter/-innen der Beschäftigten müssen deshalb nicht nur eine eigene Strategie im Umgang mit der Digitalisierung entwickeln und eine dazu passende Organisationsform einrichten. Sie müssten sich nicht selten auch darum kümmern, dazu passende Strukturen im Management durchzusetzen.

Eine zerstückelte Projektlandschaft in Unternehmen und mangelnde Informationen sind nicht die einzigen Schwierigkeiten, vor denen Vertreter/-innen der Beschäftigteninteressen stehen, wenn sie für die Arbeit mit neuen Techniken ein Regelwerk entwerfen und durchsetzen wollen (vgl. Matuschek/Kleemann 2018). Sie wollen ja etwas gestalten, dass sie noch nicht oder kaum kennen: Welche neue Technik wird eingesetzt und welche Folgen wird sie haben? Was muss genau reguliert werden? Welche Regeln werden funktionieren und welche nicht? Betriebsräte wissen meist viel zu wenig über die große Vielfalt, in der

digitalisierte Technik bereits eingesetzt wird. Und: Die technische Modernisierung ist meist viel schneller als deren betriebliche Regulierung auf der Grundlage von Betriebsvereinbarungen. Eine Lösung bieten *Prozessvereinbarungen*, welche die Rechte der Betriebsräte und der Beschäftigten in einem von Beginn an offenen Gestaltungsprozess sichern.

5.2 Perspektiven komplementärer Systemgestaltung: Managt der Mensch die Autonomen Systeme?

Die Leistungsfähigkeit Autonomer Systeme ist begrenzt. Das könnte dazu führen, dass Unternehmen die Rolle menschlicher Arbeit, auch auf lange Sicht, aufwerten. Ein entsprechendes Ziel wird im sozio-technischen Systemansatz genannt: das Ziel der *Komplementarität*. Demzufolge sollen sich Mensch und Technik zu einem Team ergänzen. Die Umsetzung wird so beschrieben:

„Statt Funktionen aufgrund vermuteter Leistungsvorteile jeweils dem Menschen oder der Technik zuzuteilen, sollte die Interaktion zwischen Mensch und Technik bei der Erfüllung der Aufgabe des Gesamtsystems unter explizitem Bezug auf ihre sich ergänzende Unterschiedlichkeit optimiert werden.“ (Grote et al. 1999: 21)

Es ist also das Ziel, eine „situationsangepasste Flexibilität“ (ebd. 21) einzurichten, sodass mal die Technik und mal die menschliche Arbeitskraft den Prozess beherrscht.

Menschen werden für Aufgaben zuständig sein, welche die Technik allein nicht zuverlässig ausführen kann. Autonome Systeme können wiederum Mittel auswählen, um vorgegebene Zwecke zu erreichen. Sie können darüber hinaus konkrete Zwecke wählen, allerdings nur in dem Rahmen, den Menschen mit ihren übergeordneten Zielen abgesteckt haben. Übliche Zwecke sind „Priorisierung, Reihenfolge, Intensität, Taktung, Dauer der Zweckverfolgung, Direktheit oder Umweg der Realisierung, bis hin zum Abbruch von Aktionen“ (Hubig

2018: 17). Das System legt weder allgemeine noch übergeordnete Ziele fest. Deshalb kann es Hubig (ebd.) zufolge weder für die Ziele noch für die Maßnahmen Verantwortung tragen, mit denen die Ziele erreicht werden sollen. Menschen müssen den Einsatz der Technik so gestalten, dass Risiken vermieden werden und ihr Einsatz ethisch vertretbar ist.

Darüber hinaus ist menschliche Arbeit aus zwei weiteren Gründen gefordert: Die Systeme sind nur unzureichend fähig, in unbekannten Situationen zu reagieren. Und sie übermitteln immer wieder zweifelhafte Informationen. Und: Werden Prioritäten verändert, dann kann es notwendig sein, von der automatisierten Routine abzuweichen. Menschliche Eingriffe in die Technik sind auch sinnvoll, wenn es um direkte Kommunikation mit Kund/-innen geht. Oder wenn zu erwarten ist, dass ein Beschäftigter/eine Beschäftigte mit seinem/ihrer Erfahrungswissen genauer und früher weiß, wann eine Wartung ansteht oder ob die Kapazitätsplanung geändert werden muss.

Aktuell wird erforscht, wie Mensch und Technik optimal zusammenarbeiten können. Ausgangspunkt ist die Überzeugung, dass Mensch und Computer zusammen leistungsfähiger sind als Mensch oder Computer alleine (vgl. Schmidt/Herrmann 2017). Im Zusammenspiel mit neuer Technik werden Menschen deshalb nicht ersetzt, sondern sie erhalten höherwertige Aufgaben. Die Forschung geht davon aus, es sei sinnvoll und im Interesse einer effizienten Produktion notwendig, *Interventionen* des Menschen nicht nur zuzulassen, sondern mit einer entsprechenden Prozessgestaltung sogar optimal zu unterstützen. Technische Systeme sollen erkennen, wann eine Intervention des Personals erforderlich ist. Darüber hinaus soll die Technik lernen, Absichten des Personals zu verstehen, ohne dass Menschen detailliert und kleinteilig in die Prozesssteuerung eingreifen müssen. Weitere Forschungsfragen lauten: Wie müssen die Informationen für eine optimale Zusammenarbeit aufbereitet werden? Und wie hat der Mensch sichere Möglichkeiten, bei Korrekturbedarf in die Steuerung automatisierter Prozesse einzugreifen? Wie lassen sich solche Interventionen mit einem Übergang in die spätere automatisierte Phase vereinbaren?

Diese Forschung passt zu einem Szenario, in dem der Mensch als ‚Manager‘ der Autonomen Systeme gilt. Das System arbeitet autonom, aber es gibt von Zeit zu Zeit die Kontrolle an den Menschen zurück. Damit besteht die Möglichkeit für Ad-hoc-Interventionen. Eine Gestaltungsrichtlinie lautet: „The user should feel like they are the one who controls the system.“ (Ebd.: 45) In diese Richtung entwickelt sich beispielsweise die Arbeit von Pilot/-innen (vgl. Weyer 2015).

So weit die Agenda der Forschung. Die Wirklichkeit entspricht jedoch bei weitem noch nicht diesem Ideal der Komplementarität. Studien zeigen, dass die Ironie der Automatisierung, wie bereits vor mehr als 30 Jahren von Lisanne Bainbridge beschrieben, immer noch besteht. Vor allem: Es gibt nach wie vor zu wenige arbeitsorganisatorische Maßnahmen, um sie zu überwinden. „Part of the underlying problem lies in the fact that system developers are still not taking appropriate account of the people that will ultimately use their systems.“ (Baxter et al. 2012: 70) Die Fähigkeit von automatisierten Arbeitssystemen, Störungen zu vermeiden und sie zu managen, hängt aber stark von den Menschen ab, die in diesen Systemen arbeiten.

Arbeitssysteme nach dem Prinzip der Komplementarität zu gestalten, ist für die Zukunft eine realistische Perspektive. Das gilt aber nicht für alle Beschäftigten. Es gibt auch Arbeiten, die überwiegend aus planerisch beherrschbarer und kontrollierbarer Routine bestehen. Hier wird sich bewahrheiten, was Shoshana Zuboff (2013: o.S.) als eines von drei Gesetzen der Digitalisierung formuliert hat: „Jede Technologie, die zum Zwecke der Überwachung und Kontrolle kolonisiert werden kann, wird, was auch immer ihr ursprünglicher Zweck war, zum Zwecke der Überwachung und Kontrolle kolonisiert.“

6. FAZIT

Vorliegende Studien stützen die Annahme, dass der Einsatz digitaler Techniken zu einem erheblichen Formwandel von Arbeit und einer großen Dynamik auf den Arbeitsmärkten führen wird. Seit Jahren

zeichnen sich fortschreitend polarisierte Arbeits- und Lebensbedingungen ab. Niemand kann die Zukunft der Arbeit mit KI und Autonomen Systemen exakt vorhersagen. Entscheidend ist jedoch: Menschen können diese Zukunft gestalten, die Technik gibt sie nicht vor.

Im Interesse guter Arbeitsbedingungen haben *betriebliche Arbeitsgestalter/-innen* die Aufgabe, Technik so auszuwählen, einzusetzen und in Verbindung mit der Arbeitsorganisation zu gestalten, dass für die Beschäftigten der größte Nutzen erreicht wird. Ein weiteres zentrales Handlungsfeld ist die Qualifizierung: Nur wer qualifiziert ist, kann an einer digitalen Arbeitswelt teilhaben. Da Automatisierung immer auch Arbeit ersetzt, bedarf es zudem einer Schutzpolitik für die möglichen Verlierer/-innen der Modernisierung. Zwar liegt der beste Schutz in der Qualifizierung, doch sie ist kein Allheilmittel. In den Unternehmen muss es begleitend eine vorausschauende Personalplanung und eine Beschäftigungssicherung geben, auf gesellschaftlicher Ebene eine gute Arbeitsmarkt- und Sozialpolitik.

Für die meisten Beschäftigten wird Arbeit aller Voraussicht nach anspruchsvoller. Mensch und Technik werden zusammenarbeiten, und die Aufgabe von Menschen wird zunehmend darin bestehen, steuernd und korrigierend in die technischen Abläufe einzugreifen. Die Digitalisierung führt also für die Mehrheit der Beschäftigten keineswegs zu mehr Fremdbestimmung und zu weniger Freiheitspielraum bei der Arbeit. Es wird allerdings, vor allem außerhalb der Industrie, zunehmend fremdbestimmte und prekäre Arbeit geben. Negative Vorbilder dafür sind beispielsweise die Fahrer/-innen von Uber, die Kommissionierer/-innen von Amazon oder Clickworker/-innen, die Kleidung in Online-shops zuordnen.

Arbeitspolitik wird sich mehr als in der Vergangenheit mit *überbetrieblichen Prozessen* befassen, denn Rationalisierung spielt sich zunehmend im Rahmen von Wertschöpfungsketten ab. Der digitale Wandel wird sich nicht auf den isolierten Betrieb und unternehmerische Rationalisierungsprogramme beschränken. Darüber hinaus entstehen mächtige Plattformen, die mit ihrer Vertragsgestaltung und Leistungs-

steuerung zunehmend Schutzstandards und geltendes Arbeitsrecht aushebeln. Es ist eine große Herausforderung für die Gewerkschaften, in dieser Situation die Vereinzelung der Beschäftigten zu überwinden und Solidarität herzustellen.

Negative Seiten der Digitalisierung zeigen sich derzeit auch jenseits und unabhängig von Arbeitsbeziehungen. Es entsteht ein „überwachender Kapitalismus“ (ebd.: o.S.), der die Kund/-innen beherrscht, und sie und ihre Daten ungefragt und weitgehend unbehelligt von staatlichen Regeln in Handelsware transformiert. Eine überforderte wirkende Politik (vgl. Wagner 2017) überlässt ganze Gesellschaften dem Treiben von Händlern der Aufmerksamkeit (vgl. Wu 2016), den privatwirtschaftlich organisierten sogenannten Sozialen Medien und den Verbreitern von alternativen Fakten. Was hier geschieht, hat nur scheinbar keine Beziehung zur Modernisierung der Arbeit in Produktion und Dienstleistung. Denn Beschäftigte werden in der Gesellschaft zu Personen geformt. Wenn an dieser Bewusstseinsbildung zunehmend private Informationsdienste und Soziale Medien beteiligt sind, vermindert dies unter Umständen die Bereitschaft von Beschäftigten, demokratische Umgangsformen im Betrieb zu unterstützen.

LITERATUR

- Arntz, Melanie/Gregory, Terry/Zierahn, Ulrich (2018): Digitalisierung und die Zukunft der Arbeit: Makroökonomische Auswirkungen auf Beschäftigung, Arbeitslosigkeit und Löhne von morgen, Mannheim: ZEW Mannheim, <ftp.zew.de/pub/zew-docs/gutachten/DigitalisierungundZukunftderArbeit2018.pdf> vom 20.8.2018.
- Bainbridge, Lisanne (1983): „Ironies of Automation“, in: *Automatica* 19/6, S. 775-779.
- Baxter, Gordon/Rooksby, John/Wand, Yuanzhi/Khajeh-Hosseni, Ali (2012): „The Ironies of Automation ... Still Going Strong at 30?“, in: Phil Turner/Susan Turner (Hg.), *European Conference on Cognitive Ergonomics, ECCE 2012, Edinburgh, 28.-31.8.2012*, New

- York: ACM, S. 65-71, <https://dblp.uni-trier.de/db/conf/ecce/ecce2012.html> vom 21.8.2018.
- Böhle, F. (1998): „Technik und Arbeit – neue Antworten auf ‚alte‘ Fragen“, in: Soziale Welt 3, S. 233-252.
- Böhle, Fritz/Stöger, Ursula/Wehrich, Margit (2015): Interaktionsarbeit gestalten. Vorschläge und Perspektiven für humane Dienstleistungsarbeit. Berlin: edition sigma.
- Böhle, Fritz/Bolte, Annegret/Neumer, Judith/Pfeiffer, Sabine/Porschen, Stephanie/Ritter, Tobias/Sauer, Stefan, Wühr, Daniela (2011): Subjektivierendes Arbeitshandeln – „Nice to have“ oder ein gesellschaftskritischer Blick auf „das Andere“ der Verwertung?, in: AIS Studien 4/2, S. 16-26.
- Braverman, Harry (1974): Labor and Monopoly Capital. The Degradation of Work in the Twentieth Century, New York: Monthly Review Press.
- Brynjolfsson, Erik/McAfee, Andrew (2014): The Second Machine Age. Work, Progress, and Prosperity in a Time of Brilliant Technologies, New York: W. W. Norton & Company.
- Broy, Manfred (Hg.) (2010): Cyber physical systems. Innovation durch softwareintensive eingebettete Systeme, Berlin: Springer.
- Butollo, Florian/Ehrlich, Martin/Engel, Thomas (2017): „Amazonisierung der Industriearbeit. Industrie 4.0, Intralogistik und die Veränderung der Arbeitsverhältnisse in einem Montageunternehmen in der Automobilindustrie“, in: Arbeit 26/1, S. 33-59.
- Carr, Nicholas (2014): „Die Herrschaft der Maschinen“, in: Blätter für deutsche und internationale Politik 2, S. 45-55.
- Castoriadis, Cornelius (1997): Gesellschaft als imaginäre Institution. Entwurf einer politischen Philosophie, Berlin: Suhrkamp.
- Dauth, Wolfgang/Findeisen, Sebastian/Südekum, Jens/Wößner, Nicole (2017): German Robots – The Impact of Industrial Robots on Workers (=IAB-Discussion Paper 30), <http://doku.iab.de/discussionpapers/2017/dp3017.pdf> vom 20.8.2018.
- Frey, Carl B./Osborne, Michael A. (2013): The Future of Employment: How Susceptible are Jobs to Computerisation? Working paper, Ox-

- ford: Oxford Martin School (OMS), https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf vom 20.8.2018.
- Gerst, Detlef (2018): „Arbeit 4.0. Den Transformationsprozess gestalten“, erscheint in: Schulz, Irene (Hg.): Industrie im Wandel. Bildungsarbeit in Bewegung. Praxishandbuch, Frankfurt a.M.: Bund Verlag.
- Goffey, Andrew (2008): „Algorithm“, in: Matthew Fuller (Hg.), Software Studies. A Lexicon, Cambridge, Mass.: MIT Press, S. 15-20.
- Gröger, Bernd (2018): Mit MES-Systemen und Kennziffern auf dem Weg zu Industrie 4.0. Aufbau und Funktionsumfang, Gestaltungs- und Regelungsansätze (= Arbeit, Gesundheit, Umwelt, Technik 86), Dortmund: Technologieberatungsstelle beim DGB NRW e.V.
- Grote, Gudela/Wäfler, Toni/Ryser, Cornelia/Weik, Steffen/Zölch, Martina/Windischer, Anna (1999): Wie sich Mensch und Technik sinnvoll ergänzen. Die Analyse automatisierter Produktionssysteme mit KOMPASS, Zürich: vdf Hochschulverlag der ETH.
- Hirschi, Caspar (2018): „Dass uns wegen der Automatisierung die Arbeit ausgeht, ist wenig plausibel“, in: Neue Zürcher Zeitung vom 3.5.2018, <https://www.nzz.ch/wirtschaft/dass-uns-wegen-der-automatisierung-die-arbeit-ausgeht-ist-wenig-plausibel-ld.1382478> vom 20.8.2018.
- Hoppe, Annette (2010): „Komplexe Technik – Hilfe oder Risiko? Darstellung ausgewählter Ergebnisse einer Grundlagenuntersuchung zu Technikstress“, in: Cornelia Brandt (Hg.), Mobile Arbeit – Gute Arbeit? Arbeitsqualität und Gestaltungsansätze bei mobiler Arbeit: ver.di – Vereinte Dienstleistungsgewerkschaft, Berlin: Gewerkschaft ver.di, S. 53-64.
- Hubig, Christoph (2018): „Soll und kann Künstliche Intelligenz ethisch handeln“, in: Institut für Geschichte und Zukunft der Arbeit (IGZA)/IG Metall (Hg.), Roboter – Assistenz-Systeme – Künstliche Intelligenz. Neue Formen der Mensch-Maschine-Interaktion, Berlin: Institut für Geschichte und Zukunft der Arbeit, S. 16-21.

- Ittermann, Peter/Niehaus, Jonathan/Hirsch-Kreinsen, Hartmut/Dregger, Johannes/ten Hompel, Michael (2016): *Social Manufacturing and Logistics. Gestaltung von Arbeit in der digitalen Produktion und Logistik* (= Soziologisches Arbeitspapier 47), Dortmund: Technische Universität Dortmund.
- Jacobs, Johann C./Kagermann, Henning/Spath, Dieter (2017): *Arbeit in der digitalen Transformation. Agilität, lebenslanges Lernen und Betriebspartner im Wandel. Ein Beitrag des Human-Resources-Kreises von acatech und der Jacobs Foundation – Forum für Personalvorstände zur Zukunft der Arbeit*, München: Herbert Utz.
- Lenzen, Manuela (2018): *Künstliche Intelligenz. Was sie ist & was sie kann*, München: C.H. Beck.
- Malsch, Thomas (1987): „Die Informatisierung des Erfahrungswissens und der ‚Imperialismus der instrumentellen Vernunft‘“, in: *Zeitschrift für Soziologie* 2, S. 77-91.
- Martini, Mario/Botta, Jonas (2018): „Iron Man am Arbeitsplatz? Exoskelette zwischen Effizienzstreben, Daten- und Gesundheitsschutz. Chancen und Risiken der Verschmelzung von Mensch und Maschine in der Industrie 4.0“, in: *Neue Zeitschrift für Arbeitsrecht* 10, S. 625-637.
- Matuschek, Ingo/Kleemann, Frank (2018): „„Was man nicht kennt, kann man nicht regeln.“ Betriebsvereinbarungen als Instrument der arbeitspolitischen Regulierung von Industrie 4.0 und Digitalisierung“, in: *WSI-Mitteilungen* 71, S. 227-234.
- Raffetseder, Eva-Maria/Schaupp, Simon/Staab, Philipp (2017): „Kybernetik und Kontrolle. Algorithmische Arbeitssteuerung und betriebliche Herrschaft“, in: *Prokla* 187/2, S. 229-247.
- Rifkin, Jeremy (2014): *Die Null Grenzkosten Gesellschaft. Das Internet der Dinge, kollaboratives Gemeingut und der Rückzug des Kapitalismus*, Frankfurt a.M./New York: Campus
- Schmidt, Albrecht/Herrmann, Thomas (2017): „User Interfaces: A New Interaction Paradigm for Automated Systems“, in: *Interactions* 25/5, Sept.-Okt., S. 41-46.

- Steinberger, Viktor (2011): „Manufacturing Execution Systems – (neue?) Aufgaben für die Belegschaftsvertretung“, in: Computer und Arbeit 8, S. 14-18.
- Timpe, Klaus-Peter/Jürgensohn, Thomas/Kolrep, Harald (2002): Mensch-Maschine-Systemtechnik: Konzepte, Modellierung, Gestaltung, Evaluation, Düsseldorf: Symposion.
- Wagner, Thomas (2017): Das Netz in unsere Hand! Vom digitalen Kapitalismus zur Datendemokratie, Köln: PapyRossa.
- Weltz, Friedrich (1999): „Der Traum von der absoluten Ordnung und die doppelte Wirklichkeit der Unternehmen“, in: Eckart Hildebrandt (Hg.), Betriebliche Sozialverfassung unter Veränderungsdruck, Berlin: Springer, S. 85-97.
- Weyer, Johannes (2006): Die Kooperation menschlicher Akteure und nicht-menschlicher Agenten. Ansatzpunkte einer Soziologie hybrider Systeme (= Arbeitspapier 16 der Wirtschafts- und Sozialwissenschaftlichen Fakultät), Dortmund: Technische Universität Dortmund.
- Weyer, Johannes (2015): Can Pilots Still Fly? Role Distribution and Hybrid Interaction in Advanced and Automated Aircraft (= Soziologisches Arbeitspapier 45), Dortmund: Technische Universität Dortmund.
- Wu, Tim (2016): The Attention Merchants. The Epic Scramble to Get Inside Our Heads, New York: Knopf.
- Zika, Gert/Helmrich, Robert/Maier, Tobias/Weber, Enzo/Wolter, Marc I. (2018): Arbeitsmarkteffekte der Digitalisierung bis 2035. Regionale Branchenstruktur spielt eine wichtige Rolle (= IAB-Kurzbericht 9), <http://doku.iab.de/kurzber/2018/kb0918.pdf> vom 20.8.2018.
- Zuboff, Shoshana (2013): „Protokoll einer Zukunftsvision. Das System versagt“, in: Frankfurter Allgemeine online vom 11.2.2013, <http://www.faz.net/aktuell/feuilleton/debatten/kapitalismus/protokoll-einer-zukunftsvision-das-system-versagt-12057446.html> vom 20.8.2018.

Assimilierende versus komplementäre Adaptivität

Grenzen (teil-)autonomer Systeme

Norbert Huchler

1. EINLEITUNG

Digitale Technologien, wie die (teil-)autonome Steuerung von Prozessen über ‚Künstliche Intelligenz‘ (KI) bzw. lernende Systeme, können höchst unterschiedlich entwickelt und eingesetzt werden. Eine generelle Folgewirkung ihres Einsatzes auf Beschäftigung, Qualifizierung, Handlungsräume, Motivation, Gesundheit etc. lässt sich nicht bestimmen. Entscheidend ist die konkrete Gestaltung der Technikentwicklung und des Technikeinsatzes in Realbedingungen. In diesem Beitrag wird argumentiert, dass sich hier zwei Vorgehensweisen unterscheiden lassen: eine assimilierende, die auf die umfassende Beherrschbarkeit durch technische Systeme setzt und in deren Eigenlogik verharret, und eine komplementäre, die Vorteile vor allem in der wechselseitigen Verschränkung technischer und humaner Koordination im Sinne einer offenen Kooperation sieht. Die Identifikation der Grenzen der Digitalisierung bietet eine Möglichkeit, um sich der zweiten Vorgehensweise, also einer neuen Arbeitsteilung zwischen Mensch und (teil-)autonomen technischem System anzunähern. Die Anerkennung von ‚Grenzen‘ bedeutet nicht, dass dadurch der digitale Wandel ins Stocken ge-

rät. Im Gegenteil: Dadurch können Innovationen angestoßen werden. Aber wenn diese Grenzen in der Gestaltung des digitalen Wandels ignoriert und nicht systematisch berücksichtigt werden, wird dies mit entsprechenden Folgen einhergehen: für Unternehmen, Beschäftigte und die Gesellschaft. Die Auseinandersetzung mit den Grenzen der Digitalisierung dient damit einer nachhaltigen Arbeits- und Technikgestaltung.

Der Beitrag sondiert zunächst (2.) unterschiedliche Entwicklungen des digitalen Wandels von Arbeit entlang der Felder Industrie 4.0, Arbeiten 4.0 und Wirtschaft 4.0. Dann wird (3.) die dezentrale Steuerung als zentrales Moment einer gemeinsamen 4.0-Kernlogik identifiziert, das Industrie 4.0, Arbeiten 4.0 und Wirtschaft 4.0 miteinander verbindet, mit dem Ziel, Adaptivität und Flexibilität im Wertschöpfungssystem zu steigern. Dieses Ziel konfligiert allerdings (4.) mit der gängigen Praxis einer planungsgeleiteten formalen technischen und organisationalen Einhegung. Doch selbst wenn (5.) diese überwunden wird, stößt die Digitalisierung und damit auch die digitale Adaptivität (teil-)autonomer Systeme an unhintergehbare Grenzen, die sich in Form von Dilemmata und Nebenfolgen (5.1), Problemen der Komplexität der physischen und sozialen Umwelt (5.2) und Nicht-Formalisierbarkeit (5.3) äußern. Diese Grenzen sind zwar ‚unhintergebar‘, können aber ‚übergangen‘ werden (5.4). In der Praxis zeigt sich (6.), dass dies oftmals geschieht mit entsprechenden Folgewirkungen. So wird häufig der vollständige Ersatz eines vormals ‚analogen‘ Prozesses durch (Teil-)Automatisierung suggeriert, während es sich ‚nur‘ um ein funktionales Äquivalent handelt, mit einer anderen Bearbeitungslogik und anderen Folgewirkungen (6.1). Durch die Ausblendung impliziten Arbeitshandelns und Wissens werden die mit der Digitalisierung verbundenen Substitutionspotenziale systematisch überschätzt (6.2). Unter anderem um Adaptivitätsdefizite auszugleichen, tendiert eine in ihrer Eigenlogik verhaftete Gestaltung (teil-)autonomer Technologie dazu, die Umwelt an die Bedarfe des Systems anzupassen (6.3). Dabei geht es oftmals eher um das Schaffen von Akzeptanz als um echte Partizipation in der Gestaltung und im Nutzungsprozess (6.4). Insgesamt dominiert also

eher ein assimilierender Zugang zu Adaptivität als ein komplementärer (6.5). Der Beitrag schließt damit, dass die Gestaltung offener sozio-technischer Systeme als Herausforderung für die Technikentwicklung insbesondere über lernende Systeme identifiziert wird (7.).

2. DIGITALER WANDEL VON ARBEIT

Im Diskurs um den digitalen Wandel von Arbeit haben sich drei Themenfelder bzw. Perspektiven herausgebildet¹, die sich – nicht völlig trennscharf – unter den Begriffen Arbeit 4.0, Industrie 4.0 und Wirtschaft 4.0 bündeln lassen. Abbildung 1 ordnet diesen drei Themenfeldern einzelne Schlagworte des Diskurses zu.

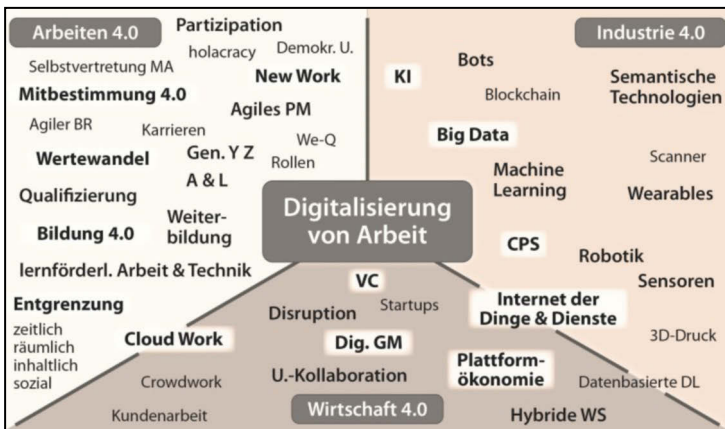
Unter dem Schlagwort *Arbeiten 4.0* geht es vor allem um die Frage, wie sich Arbeit bzw. Arbeitsstrukturen durch Digitalisierung wandeln. Im Mittelpunkt stehen bereits bestehende und durch die Digitalisierung verstärkte Wandlungstendenzen von Arbeit in Richtung einer (zeitlichen, räumlichen, inhaltlichen, sozialen) Entgrenzung, eines Generationen- und Wertewandels, angepasster Partizipations- und Mitbestimmungsformen und neuer Arbeits- und Organisationsformen wie ‚holacracy‘ (digitale Entscheidungseinbindung) und Agilität.

Als Kern von *Industrie 4.0* lässt sich hingegen die Idee der ‚Cyber-Physical Systems‘ (CPS) ausmachen, also die (teil-)autonome Steuerung von miteinander vernetzten bzw. ‚kommunizierenden‘ Objekten über Software auf Basis von Sensordaten (Big Data) und ‚Künstlicher Intelligenz‘. Dies betrifft sowohl die Anlagen- und Prozesssteuerung in der Produktion als auch IT-Systeme in der Sachbearbeitung und an der Kundenschnittstelle.

1 Insbesondere in Folge der politischen Begleitung durch die Ministerien: Bundesministerium für Arbeit und Soziales (BMAS), Bundesministerium für Bildung und Forschung (BMBF), Bundesministerium für Wirtschaft und Energie (BMWi).

Wirtschaft 4.0 dreht sich vordringlich um die Koordination von Wertschöpfung über digitale Plattformen, die sich zwischen Produkten bzw. Anbietern und dem Markt bzw. Nachfragern positionieren, z.B. für die Vermittlung von Arbeitskräften und Ideen (Cloudwork), Dienstleistungen, Prozessdaten, Kontakten etc. Im Fokus stehen dabei neue digitale Geschäftsmodelle im Internet der Dinge und Dienste.

Abbildung 1: Digitalisierung von Arbeit – Umsetzungsfelder



Quelle: eigene Darstellung

Mit Blick auf reale Umsetzungen lässt sich feststellen, dass für den digitalen Wandel allgemein bzw. für die einzelnen Technologien und Maßnahmen keine determinierten Auswirkungen auf die Arbeit zu identifizieren sind. Sie führen beispielsweise nicht zwangsläufig zu mehr oder weniger Beschäftigung, zu Höher- oder Niedrigqualifizierung, zu Motivation oder Demotivation, zu erweiterten Handlungsräumen oder restriktiverer Arbeitsgestaltung. Es kommt vielmehr darauf an, wie der entsprechende Technologieeinsatz konkret gestaltet wird bzw. wie die Maßnahmen in das gesamte sozio-technische System im Betrieb eingebettet werden (vgl. Hirsch-Kreinsen/Ittermann/Niehaus 2015; Sydow 1985).

So kann der Einsatz von ‚Wearables‘, etwa einer Datenbrille in der Montage, so erfolgen, dass jeder Arbeitsschritt visuell im Detail kleingetaktet vorgegeben und genau kontrolliert wird und es kaum mehr Raum für eine Gestaltung der Arbeitsorganisation durch die betroffenen Beschäftigten gibt. Die gleiche Technologie kann aber auch so eingesetzt werden, dass sie, nach von den jeweiligen Beschäftigten festgelegten Intervallen, das Arbeitsergebnis auf mögliche Fehler kontrolliert, ihnen ein Feedback gibt oder auf Anfrage punktuell gezielte Informationen bzw. Hilfestellungen zu den Arbeitsgegenständen zur Verfügung stellt – ohne dass dies in die Leistungsbemessung einfließt. In diesem Fall fungiert sie als befähigendes Assistenzsystem, das Autonomiespielräume nicht beschneidet. Die Einführung einer neuen Technologie ist also immer eingebunden in ein politisches und kulturelles Gefüge, das sich in der Regel zwischen einem *Kontrollansatz* und einem *Befähigungsansatz* bewegt (vgl. Brödner 2015; Grote 2015; Huchler 2016; Huchler/Rhein 2017; Niehaus 2017).

Diese Gegenüberstellung lässt sich auf die Arbeitsgestaltung insgesamt anwenden, also auch auf Konzepte, die nur am Rande etwas mit Digitalisierung zu tun haben (im Feld Arbeiten 4.0)², oder eher auf der Ebene des Geschäftsmodells liegen (Wirtschaft 4.0)³. Angesichts der

-
- 2 Z.B. besteht bei Holacracy (vgl. Robertson 2016) die Gefahr, das Ziel ‚demokratischer‘ Strukturen durch Überformalisierung zu gefährden; oder agiles Projektmanagement, das auch als kleinteiliges Kontrollinstrument der Leistungssteuerung von vormalig selbstorganisierter Wissensarbeit verwendet werden kann (vgl. Neumer/Porschen-Hueck/Sauer 2018). Auch bei Home Office und Vertrauensarbeitszeit können potenzielle Freiräume durch ‚indirekte‘ Steuerungsmethoden (vgl. Huchler/Voß/Weihrich 2007; Moldaschl/Sauer 2000; Moldaschl/Voß 2003) in ein enges Kontrollkorsett eingebettet und so konterkariert werden.
 - 3 Die Organisation von Teilen des Wertschöpfungsprozesses über digitale Plattformen kann dazu genutzt werden, Prozesse top-down zu verschlanken und Kosten einzusparen bzw. Arbeit an die Kund/-innen auszulagern, oder auch dafür, das Geschäftsmodell aus den Prozessen heraus und partizipativ

Heterogenität der Konzepte, der Breite ihrer Umsetzungsmöglichkeiten und der Fortschreibung typischer Problematiken der Arbeitsgestaltung (vgl. Huchler/Pfeiffer 2018) muss gefragt werden, was eigentlich das Gemeinsame und Neue ist, das diese 4.0-Konzepte miteinander verbindet. Hier reicht der Verweis auf Digitalisierung allgemein und die Kennzeichnung einer ‚vierten industriellen Revolution‘ aufgrund von ‚intelligenter‘ Vernetzung (vgl. Promotorengruppe Kommunikation der Forschungsunion Wirtschaft – Wissenschaft 2013) allein nicht aus.

3. DIE 4.0-LOGIK: DEZENTRALE STEUERUNG UND ADAPTIVITÄT

Was hält das vielfältige Setting von Perspektiven und Gegenständen der 4.0-Diskussion zusammen? Was macht 4.0 eigentlich aus? Die 4.0-Thematik wird einerseits angetrieben von bekannten Rationalisierungs-ideen, wie Einsparung von Personalkosten und Unabhängigkeit von Arbeitskraft durch Automatisierung, verbunden mit dem Abzielen auf neue Märkte und Geschäftsmodelle (vgl. Pfeiffer 2015: 16). Andererseits liegt ihr auch eine Zeitdiagnose zugrunde, die explizit auf etwas Neues zielt. Sie lässt sich mit dem Begriff VUKA umschreiben: VUKA steht für *Volatilität*, insbesondere schnelle Marktschwankungen bzw. hohe Marktsensibilität, zunehmende *Unsicherheit* und Unplanbarkeit, hohe *Komplexität*, u.a. durch zunehmend vernetzte Prozesse, und häufigere *Ambiguität*, also ein Sowohl-als-auch, das eindeutige Zuordnungen weniger möglich macht.

Vor dem Hintergrund von VUKA stoßen klassische Formen formaler Arbeits-/Prozessplanung an Grenzen. Die Flexibilitätsdefizite der Kombination aus *Top-down-Steuerung* und *formaler Vorausplanung* über stabile Prozesse und Regeln werden offensichtlich: in Bezug auf

mit Beschäftigten und Kund/-innen innovativ zu erweitern, neue Felder für die Zukunft zu erschließen oder Marktanteile auszubauen.

hierarchische und bürokratische Arbeitsorganisation (Arbeiten 4.0), die lineare und feste Gestaltung von Produktionsprozessen (Industrie 4.0) und die Stabilität von Wertschöpfungsketten (Wirtschaft 4.0). Damit werden diese Formen der Organisation nicht obsolet, sie bieten jedoch für sich allein nicht mehr konkurrenzfähige Lösungen für die Flexibilitätsbedarfe der VUKA-Welt. Das ergänzende neue Lösungsprinzip 4.0 lautet: *dezentrale Steuerung und Adaptivität*.

Im Zentrum stehen die *Vernetzung* von Prozessen und deren *dezentrale Koordination* von unten aus dem Prozess heraus: bei Arbeiten 4.0 durch die Mitarbeiter/-innen bzw. aus deren *Arbeitshandeln* heraus, bei Industrie 4.0 vom Produkt her bzw. aus dem *Produktionsprozess* heraus und bei Wirtschaft 4.0 vom *Nutzungsprozess* und den dort anfallenden Daten aus.

Die Prozesse, die im Kontext von Arbeiten 4.0 erneut aufgegriffen und mit Digitalisierung verknüpft werden, lassen sich im Kern mit dem arbeitssoziologischen Begriff der Subjektivierung von Arbeit (vgl. Moldaschl/Sauer 2000; Moldaschl/Voß 2003) fassen. Im Mittelpunkt steht der zunehmende Einbezug des ‚ganzen Subjekts‘ in den Verwertungsprozess. Dabei geht es über die formalen Qualifikationen hinaus um Kompetenzen und grundlegende Fähigkeiten, wie Kreativität, Empathie und Initiative, aber auch um informelle Leistungen (vgl. Huchler/Voß/Wehrich 2007: 62ff.), implizites Wissen und Erfahrungswissen (vgl. Böhle/Porschen 2012; Böhle/Rose 1992), das vor Ort generiert und genutzt wird, also am Kunden, am Gegenstand, im Prozess. Schon im Prozess werden Lösungen erarbeitet und Innovationen umgesetzt. Teilweise wird die Arbeitsorganisation aus dem Prozess heraus verändert. Widersprüche, die nicht formal geregelt werden können (z.B. in Sinne eines Grenznutzens), werden im Prozess austariert und ausgetragen, etwa zwischen den ökonomischen Zielvorgaben und Qualitätsversprechen gegenüber Kund/-innen (vgl. Huchler/Voß/Wehrich 2007: 158, 278ff.). Insgesamt geht es dabei um die Kompetenz zur Selbstorganisation.

Industrie 4.0 steht für die Übertragung der Logik der Selbstorganisation auf technische Systeme: die Steuerung vom Produkt her bzw.

aus dem Produktionsprozess heraus. Die Defizite bisheriger Prozessplanung und -steuerung, wie die eingeschränkte Fähigkeit zur Reaktion auf veränderte Anforderungen (individualisierte Produkte, kleine Losgrößen, Auftragsschwankungen) und die zum Teil hohe Distanz von aktuell laufenden Prozessen, sollen mittels „dezentraler“ (Promotorengruppe Kommunikation der Forschungsunion Wirtschaft – Wissenschaft 2013: 56f., 95ff.) und „intelligenter“ (Plattform Industrie 4.0 2015: 28f.) bzw. lernender Systeme überwunden werden. Die Überlegung ist, dass vor Ort vorhandene Daten vertikal und horizontal im Wertschöpfungssystem vernetzt und integriert werden, um damit eine situative, flexible und integrierte bzw. ‚ganzheitliche‘ Steuerung in Echtzeit zu ermöglichen (vgl. Plattform Industrie 4.0 2015, 73f.; Promotorengruppe Kommunikation der Forschungsunion Wirtschaft – Wissenschaft 2013: 5f., 28f.). Mit neuen Methoden der Datenanalyse und des ‚Machine Learning‘ sollen Vorhersagen generiert und Entscheidungen getroffen werden, ein großer Teil der Steuerung soll also an das autonom entscheidende System übergeben werden. Klassische Beispiele aus Modell-Fabriken sind Produktionsverfahren, die von den einzelnen Produkten aus teilgesteuert werden, wie im Fall von RFID-Chips auf Flaschen, die selbsttätig Befüllung und Etikett anfordern.

Im Zentrum der Wirtschaft 4.0 geht es um eine plattformgesteuerte Wertschöpfung, bei der die Vertriebslogik die Herstellungslogik dominiert. Die Steuerungslogik fokussiert auf den Absatz und die Daten, die beim Nutzungsprozess anfallen (ggf. zum Kundennutzen bzw. ‚Kundenerlebnis‘).

Insgesamt ist es das Ziel, Situativität, Innovativität, Bedarfsorientierung, Flexibilität und Effizienz zu erhöhen, um so innerorganisatorische und unternehmensübergreifende Wertschöpfungsprozesse zu verschlanken, zu verdichten und zu beschleunigen, aber auch zu flexibilisieren.

Die Strategie der (punktuellen) Verlegung auf *dezentrale Steuerung* ist nicht neu, weder in Bezug auf Konzepte der Arbeitsorganisation noch auf Technikleitbilder. Sie gewinnt aber mit zunehmender (digita-

ler) Vernetzung und breiterer Anwendung von lernenden Systemen an diskursiver Brisanz und Umsetzungsrelevanz.

4. DIGITALE ADAPTIVITÄT VERSUS FORMALE PLANUNG

Die Idee der *Dezentralisierung* zur Ausschöpfung von Rationalisierungspotenzialen gewinnt also wieder an Bedeutung. Die Flexibilitäts- und Innovationsdefizite von hierarchischer Top-down-Steuerung, Vorausplanung, Zentralisierung in einzelnen abgegrenzten Zuständigkeitsbereichen und linear angelegten Prozessen sollen erneut angegangen werden. Dabei wird nunmehr jedoch auf dezentrale, vernetzte technische (teil-)autonome Systeme bzw. „Multiagentensysteme“ und dezentrale sozio-technische Systeme (also eine neue Mensch-Technik-Kollaboration, z.B. mittels „multimodaler Assistenzsysteme“ [Plattform Industrie 4.0 2015: 29]) gesetzt. Entsprechend dem Leitbild des Internets der Dinge (IoT) sollen Flexibilitätserhöhung, Verschlankeung, Dynamisierung und Produktivitätssteigerung durch dezentrale, ganzheitlich vernetzte, integrierte, komplexe Wertschöpfungssysteme erreicht werden, die – über digitale Plattformen vermittelt – sich in weiten Teilen selbst ‚von unten‘ dezentral steuern.

Eine notwendige Voraussetzung hierfür ist, dass die Systeme ‚adaptiv‘ sind, also in der Lage sind, sich veränderten Umweltbedingungen selbsttätig anzupassen. Dies soll auf Basis von Daten vor Ort, vom Objekt aus, aus dem Prozess heraus, aus der Anwendung oder von dem Nutzer/der Nutzerin aus geschehen, die durch sich selbst optimierende, lernende KI-Systeme in Echtzeit ausgewertet werden, welche hierauf wiederum neue – bislang ungeplante – Handlungen ‚autonom‘ anstoßen (vgl. exemplarisch Plattform Industrie 4.0 2015; Promotorengruppe Kommunikation der Forschungsunion Wirtschaft – Wissenschaft 2013).

Eine wesentliche Rolle spielen dabei statistische Methoden der Auswertung von Daten (Big Data), Mustererkennung, semantische

Technologien, Deep Learning etc. (vgl. Promotorengruppe Kommunikation der Forschungsunion Wirtschaft – Wissenschaft 2013). Die Kombination von Hard- und Software zu Cyber-Physical Systems ist hierfür die Ausgangsbasis. Die Entwicklung hin zu CPS geht mit einer ‚Informatisierung‘ (vgl. Baukrowitz et al. 2006; Boes 2005) einher, mit der Transformation von Wissen und Prozessen ins Digitale. Dahinter steht die Idee eines „Echtzeitabbilds“ (Promotorengruppe Kommunikation der Forschungsunion Wirtschaft – Wissenschaft 2013: 108), „digitalen Abbilds“ (Plattform Industrie 4.0 2015: 23) oder auch ‚digitalen Zwillings‘ in Form einer ‚strukturellen Dopplung‘ der analogen Welt in der digitalen (siehe auch den Beitrag von Thomas Kuhn und Peter Liggesmeyer in diesem Band, S. 41ff.). Diese soll über den ‚digitalen Raum‘ bzw. den „Informationsraum“ (Boes 2005: 222ff.) den direkten synchronisierten, raumunabhängigen Zugriff in Echtzeit auf möglichst alle analogen Sachverhalte ermöglichen, bei größtmöglicher Transparenz, um die Arbeits- bzw. Wertschöpfungsprozesse zu analysieren und zu optimieren. Damit werden im Sinne eines Social Engineerings technikorientierte Überlegungen auf die Organisation übertragen. Weitergedacht entsteht so, gekoppelt mit KI und Machine Learning, ein sich selbst optimierendes und selbst steuerndes organisationales System (vgl. Huchler/Rhein 2017). Wie weit reicht nun diese digitale Dezentralisierung? Wo und wie verbindet sie sich mit der parallel weiter existierenden planungsgeleiteten Unternehmenssteuerung?

Die Formalisierung bzw. die Hinterlegung von Prozessen durch deren Verschriftlichung (Daten) ist der wesentliche Schalthebel zur Objektivierung, Explizierung und Messbarmachung (vgl. Böhle et al. 2011). Dies ist einerseits die Voraussetzung für adaptive (teil-)autonome Systeme, z.B. indem Sensoren Bilder, Töne, Texte, Berührungen etc. in verarbeitbare Daten ‚übersetzen‘ und auf deren Basis ‚agieren‘. Andererseits ermöglichen diese Daten in aufbereiteter Form auch erhöhte Planbarkeit und Steuerbarkeit (z.B. über Eingriffe in die Systemarchitektur), Transparenz, Dokumentation, Reproduzierbarkeit und Anschlussfähigkeit für andere formale (z.B. ökonomische, rechtliche und technische) Systeme. So wird in Anwendungsbeispielen die

Kernidee dezentraler sozio-technischer Systeme oft verknüpft mit der Vorstellung einer integrierten bzw. ‚zentralisierten‘ Steuerung und ‚Durchsteuerbarkeit‘ von komplexen Wertschöpfungssystemen und Organisationen – über die Systemarchitektur und mittels aktueller Daten. Beliebte ist die Idee eines ‚Cockpits‘ (in der Regel ein Tablet), mit dessen Hilfe komplexe Systeme (z.B. die Produktion) mittels technisch ‚vorbereiteter Entscheidungen‘ (vgl. Mühge 2018) gesteuert werden können.

(Teil-)Autonome Systeme bleiben in ein von hierarchischen und linearen Prozessen dominiertes betriebliches Umfeld eingebettet, ebenso wie das bei den Mitarbeiter/-innen in offenen Arbeitsorganisationen (Gruppenarbeit usw.) der Fall ist. Dieses Umfeld ist insbesondere durch die Systemarchitektur geprägt, aber auch durch angrenzende Systeme und Prozesse, Vorgaben, Regelungen und Eingriffe. Das typische betriebliche Wechselspiel zwischen der Kontroll- bzw. Beherrschungslogik mit dem Ziel einer möglichst detaillierten, planmäßigen Ex-ante-Steuerung und Messbarkeit der Resultate einerseits und systematischer Angewiesenheit auf die Selbststeuerung in nicht beherrschbaren Autonomieräumen vor Ort andererseits lässt sich auch auf (teil-)autonome Systeme anwenden. Entsprechend geht es aktuell um den Versuch, möglichst große Anteile des Wertschöpfungsprozesses dezentral zu organisieren, nicht nur durch mehr Selbstorganisation durch die Mitarbeiter/-innen, sondern parallel und oftmals in Konkurrenz dazu auch durch (teil-)autonome bzw. lernende Systeme. Neue Technologien bieten potenziell die Möglichkeit, einen weiteren Schritt in Richtung Dezentralität zu gehen.

Das Wechsel- und Zusammenspiel von dezentralen (teil-)autonomen Systemen und linearer Unternehmenssteuerung bleibt eine offene Gestaltungsfrage. Es stellt sich aber eine darüber hinausreichende grundlegende Frage: Wie weit sind adaptive Systeme überhaupt in der Lage, Aufgaben (teil-)autonom zu übernehmen? Auf welche Aufgaben trifft das zu? Wo liegen ihre Grenzen?

5. DIE GRENZEN DIGITALER ADAPTIVITÄT DURCH LERNENDE SYSTEME

Die gestiegene Vielfalt an Sensoren, die steigende Menge an vorhandenen Daten, die zunehmende Rechenleistung und teilweise auch die neuen Rechenmethoden prägen die weit verbreitete Idee von ‚selbstständig‘ agierenden KI-Systemen. Kaum eine öffentliche Diskussion oder Tagung zum Thema kommt ohne Fragen zur neuen Kooperation mit dem ‚Kollegen KI/Roboter‘, zum Substitutionspotenzial, der verbleibenden Rolle des Menschen, den Unterschieden zwischen Mensch und lernendem technischem System und zur nötigen rechtlichen Rahmung von eigenständigen Systemen aus.

Als entscheidender ‚evolutionärer‘ Schritt wird der Wandel weg vom passiven Objekt hin zu einem ‚intelligenten‘ und schließlich lernenden System gesehen, das selbsttätig mit seiner Umwelt ‚interagieren‘ kann – also letztlich adaptiv ist. Lernende Systeme können Muster bzw. statistische Regelmäßigkeiten erkennen und Wahrscheinlichkeiten abgleichen, um daraus ‚Handlungsannahmen‘ abzuleiten.⁴ Auf Grundlage vorgegebener Handlungsmodelle können diese technischen Systeme selbst ‚Entscheidungen‘ treffen und Aktionen anstoßen. So können sie statistisch ‚antizipieren‘ bzw. einordnen und vorhersagen, wie sich Menschen in offenen Situationen und bei unterschiedlichen Handlungsmöglichkeiten entscheiden werden, und dementsprechend handeln (vgl. Fink/Weyer 2014).

Verschiedene soziologische Theorieperspektiven haben die traditionelle Gegenüberstellung von menschlichem Akteur und passivem technischem Objekt bereits grundlegend aufgebrochen und technischen Artefakten die Rolle eines Akteurs bzw. Aktanten zugesprochen (vgl. Rammert/Schulz-Schaeffer 2002; vgl. auch Latour 1987). Solche Über-

4 Wie Big-Data-Algorithmen oder die KI ‚Watson‘ von IBM mit ihrer DeepQA bzw. „massively parallel probabilistic evidence-based architecture“ (www.watson2016.com vom 28.9.2018).

legungen zu einer Neuverteilung der Handlungsträgerschaft zwischen Mensch und Technik (siehe auch den Beitrag von Ingo Schulz-Schaeffer in diesem Band) gewinnen mit der aktuellen breiteren Aufmerksamkeit bezüglich lernender Systeme bzw. KI deutlich an Relevanz (vgl. Huchler 2018). Um mögliche Formen der neuen Arbeitsteilung zwischen Mensch und Technik bestimmen zu können, ist als ein wesentlicher Teilschritt die Identifikation und Systematisierung der Grenzen (teil-)autonomer und lernender Systeme notwendig (ebd.); insbesondere auch was deren Adaptivität angeht, also die Fähigkeit, sich an sich ändernde Umweltbedingungen selbsttätig anzupassen.

Im Folgenden wird ein praxisbezogener, aber theoriegeleiteter Weg gewählt, um die Grenzen zu sondieren, die sich bei der Umsetzung digitaler (Teil-)Automatisierung identifizieren lassen. Diese können – als ein Systematisierungsversuch – unterschieden werden in 1) Dilemmata und Nebenfolgen, 2) Grenzen durch die Komplexität der physischen und sozialen Umwelt und 3) Nicht-Formalisierbarkeit.

5.1 Dilemmata und Nebenfolgen

Häufig wird mit dem Begriff der ‚Automatisierungsdilemmata‘ bzw. der ‚Ironies of Automation‘ (vgl. Bainbridge 1983) darauf verwiesen, dass technische Rationalisierungsmaßnahmen selbst wieder neue Arbeit und Folgekosten erzeugen, um die Systeme einzuführen, aufrechtzuerhalten und zu optimieren, aber auch um ihre Fehler, Dysfunktionalitäten und unintendierten Folgen zu kompensieren. Zum einen müssen automatisierte Systeme begleitet werden, was mit einem zusätzlichen Organisationsaufwand einhergeht. Zum anderen sind sie in der Regel komplizierter und bestehen aus mehr Komponenten, die stärker miteinander verschränkt sind. Wenn die möglichen Interdependenzen nicht mehr planbar und in ihren Wirkungen nicht mehr eindeutig vorhersagbar sind, markiert dies den Schritt zum komplexen System. Hier versagen die klassischen Formen der Risikobeherrschung (vgl. Beck 1986), denn diese sind nicht mehr „durch ein ‚Mehr‘ an Wissenschaft und Technik ausschaltbar, sondern entstehen immer wieder in neuer Weise

und auf neuem Niveau“ (Böhle 2009: 208). Resultat sind selbstproduzierte Nebenfolgen auf ständig neuem Niveau. Dabei lagern sich zu neuen spezialisierten Organisationen um schwere Risiken herum an (vgl. Weick/Sutcliffe 2003), wie Unfälle und Katastrophen bzw. Systemausfälle, die vor dem Hintergrund der steigenden Komplexität technischer Systeme zunehmend ‚normal‘ werden (vgl. Perrow 1987). Zum anderen entstehen bereits im Kleinen ständige Störungen, Dysfunktionalitäten, Ausfälle, Änderungsbedarfe, neue Anfälligkeiten etc., die Updates, Wartungsarbeiten oder einen Austausch notwendig machen. Technische Rationalisierung erzeugt so immer neue Arbeit.

Dies zeigt sich etwa im Gespräch mit Zuständigen für die Digitalisierung der Kundenschnittstellen von Krankenkassen: Obwohl sich insgesamt Arbeitszuschnitte und -anforderungen stark ändern können und Prozesse zum Teil sehr beschleunigt werden, bleibt die Menge an Arbeit bzw. die Anzahl der notwendigen Arbeitskräfte erstaunlich konstant. So können nach einer Digitalisierung des gesamten Kommunikationsverkehrs zu Schadensmeldungen semantische Technologien eingesetzt werden, um Schadensfälle ‚im Dunkelfeld‘⁵, also im System selbsttätig zu bearbeiten. Sie können relevante Daten, wie Beteiligte, Schadenshöhe, Begründungen etc., auslesen und mit vorhandenen Daten abgleichen: Wie häufig hat die Person bereits etwas gemeldet? Lohnt sich der Aufwand einer genaueren Überprüfung? Es wird anvisiert, mittels ständiger Perfektionierung und Anpassung und nach einem entsprechend umfangreichen ‚Training‘ des lernenden Systems bis zu 80 Prozent der Fälle auf diese Weise im Hintergrund zu bearbeiten. Ein entsprechendes System muss jedoch zunächst entwickelt, eingeführt und ‚trainiert‘ werden. Hier kann zwar auf gewisse Skaleneffekte gehofft werden. Jedoch sind immer Anpassungen, Problembehebungen, Wartungen und Überprüfungen erforderlich, ob das System auch tatsächlich die geplanten Einordnungen macht. Auf dieser Seite ent-

5 Als Schatten- oder Dunkelfeldbearbeitung werden die Prozesse bezeichnet, die das System im Hintergrund selbst bearbeitet und z.B. später aufbereitet an die Sachbearbeitung oder direkt an andere Stellen/Kund/-innen übergibt.

steht also neue technische ‚Gewährleistungsarbeit‘, für die händeringend nach Fachkräften gesucht wird. Zugleich können jedoch auch die Fälle, die das System nicht bearbeiten kann oder als auffällig an Sachbearbeiter/-innen weiterreicht, nicht mehr so bearbeitet werden wie zuvor. Denn es handelt sich nun durchweg um komplexe, schwierige oder konfliktträchtige Fälle, die belastender sind, mehr Zeit in Anspruch nehmen und für die auch ein höherer Qualifikationsbedarf nötig sein kann. Da einfache Fälle fehlen, die in Routine ‚abgearbeitet‘ werden können, wird die Arbeit intensiviert, was sich auch in der Zeitspanne niederschlagen kann, über die eine solche intensivierte Fallarbeit am Stück geleistet werden kann. Dies kann Auswirkungen auf die Personalplanung haben. Nicht zuletzt müssen sich auch die Kund/-innen anpassen.

Insgesamt lassen sich folgende Prozesse in Bezug auf Beschäftigung ausmachen: Benötigt wird zum einen ein Mehr an neuer ermöglichender Arbeit (z.B. Systementwickler/-innen), an kompensierender problemlösender bzw. innovativer Arbeit vor Ort (vgl. Neumer/Pfeiffer 2015) und an gewährleistender Arbeit (Wartung, Systemanpassung etc.; z.B. Techniker/-innen). Zum anderen wird die verbleibende Arbeit intensiver und anspruchsvoller.

5.2 Grenzen durch die Komplexität der physischen und sozialen Umwelt

Je mehr Laborexperimente, Simulationen, Einzelentwicklungen und Umsetzungen in exklusiven separierten Vorzeigebereichen entstehen, desto offensichtlicher wird der Unterschied zwischen technischer Machbarkeit im Einzelnen und Implementierbarkeit in der Breite. Hier spielen nicht nur die teils enormen Investitionskosten (vgl. Schröder 2017: 11ff.), die Ressourcenproblematiken (z.B. seltene Erden) und Umweltproblematiken eine Rolle, sondern vor allem die Qualität und die Komplexität sozio-technischer Systeme. Mit Blick auf komplexe sozio-technische Systeme macht es Sinn, drei Ebenen systematisch zu unterscheiden: die Welt der Software, die Welt der Hardware bzw. der

Maschinen und die Welt des Sozialen bzw. der Arbeit. An den Übergängen sind Übersetzungsprobleme wahrscheinlich.

Von der Software zur Hardware: Die Verbindung von informativ-verarbeitenden (virtuellen) Objekten und Prozessen mit realen (physischen) Objekten und Prozessen ist eine wesentliche Herausforderung von Cyber-Physical Systems (vgl. Böhle/Huchler 2016). Während die Welt der Software zumindest im Prinzip zu 100 Prozent kontrollierbar und nach außen abschirmbar ist, Prozesse beliebig oft manipuliert und wiederholt werden können und bei Änderungsbedarfen einfach ein neues Update aufgesetzt werden kann, ist die physikalische Welt der Hardware weit weniger kontrollierbar. Die Integration in notwendigerweise bereits bestehende und oftmals nicht einfach veränderbare Systeme und die Abstimmung mit heterogenen Komponenten stellt eine enorme Herausforderung dar. Es kann zu Wechselwirkungen, Aufschaukelungseffekten, Störungen etc. kommen. Auch unabhängig von typischen Problemen der Systemintegration können aufgrund existierender physikalischer ‚Unschärfen‘ (z.B. elektrische Ladungen, Staub in der Luft) die Umwelt- und Prozessbedingungen von Hardware nicht vollständig erfasst, geplant und beeinflusst werden. Entsprechend greift man auf Tests, Messungen und Statistiken zurück. Edward A. Lee (2008) verdeutlicht diese Herausforderung vernetzt arbeitender eingebetteter CPS am Beispiel der nie vollständigen Zeitsynchronität interagierender Hardwarekomponenten. Für Lee ist die „physical world“ im Vergleich zur Cyberwelt der Software nicht vollständig vorhersagbar, denn für CPS gibt es keine kontrollierte Umwelt (vgl. ebd.: 4). Zwar sollte man nach Lee versuchen, Verlässlichkeit und Vorhersagbarkeit auch in CPS herzustellen, jedoch muss dies durch eine Strategie der Robustheit begleitet werden, da letztlich nur diese erreichbar ist (vgl. ebd.).

Vom technischen zum sozio-technischen System: Mindestens genauso gravierend ist der Schritt in die *soziale* Einbettung technischer Systeme. Ebenso wie sich die physische Welt nicht einfach mit ein paar Befehlen ändern und dann neu starten lässt, können unpassende, falsch funktionierende oder beschädigte Teile der sozialen Welt nicht

einfach ausgetauscht und – quasi als Update – an den Bedarf technischer Systeme angepasst werden. Autonomes Fahren, intelligente Häuser und Fabriken, intelligente neue Service-Systeme und Apps müssen eingebettet werden in eine bestehende dynamische, sich ständig verändernde und sehr heterogene soziale Praxis, die gerahmt ist durch individuelle und kollektive Interessen und Aneignungsprozesse, Konkurrenzen und Wettbewerb, Normen und Werte, Regeln und Gesetze, mediale Diskurse etc. So werden Ambient-Assisted-Living-Systeme (AAL) in ihrer konkreten Nutzung von älteren Menschen höchst individuell angeeignet (vgl. Birken et al. 2018), Softwareanwendungen von Nutzer/-innen und Nutzergruppen unprogrammiert und Systeme im öffentlichen Raum manipuliert (z.B. Kaugummis auf Sensoren von autonomen Fahrzeugen geklebt).

Mit jedem dieser beiden Schritte (Software – Hardware – soziale Einbettung) tut sich neue, zusätzliche Komplexität auf, die systematisch die Erfassbarkeit durch digitale Systeme überschreitet. Die Komplexität sozio-technischer Systeme stellt eine Art ‚natürliche‘ Grenze der Digitalisierung dar (vgl. Hirsch-Kreinsen/Ittermann/Niehaus 2015; Sydow 1985). Das gilt auch für die digitale Adaptivität durch (teil-)autonome lernende Systeme; auch sie blendet einen erheblichen Anteil der Komplexität sozialer Praxis aus.

Insgesamt zeigen sich im Fortgang von der kontrollierten Laborsituation zur Umsetzung in der Breite in drei Feldern Problematiken: bei *Kosten und Ressourcen*, bei der technischen Erfass- und Verarbeitbarkeit *physischer Komplexität* und bei der *sozialen Einbettung*. Diese stellen ‚materielle‘ Quellen für Grenzen digitaler (teil-)autonomer Systeme dar, also Grenzen, die sich aus der Konfrontation mit den objektiven Gegebenheiten einer komplexen Umwelt ergeben.

5.3 Nicht-Formalisierbarkeit

Die Suche nach dem Nicht-Digitalisierbaren erfordert eine Unterscheidung zwischen dem, was durch Technik *noch nicht* übernommen werden kann, und dem, was grundsätzlich *nicht* in die Logik der Technik

übersetzt werden kann. Einen Schlüssel zu dieser Unterscheidung bietet der Blick auf die formale ‚Zeichenlogik‘ (vgl. Rammert 2009) digitaler Systeme. Durch Digitalisierung wird *Wissen* in explizite bzw. ‚greifbare‘ *Information* gewandelt, also in die ‚Zeichensprache‘ (ebd.) der Informationstechnik bzw. der Software übertragen. Wissen wird hierdurch formalisiert und damit erst wirklich zugänglich gemacht für die Regelwelt von Organisationen und die Rechenlogik von Maschinen, wie definierte Prozesse, Messung und Kontrolle.⁶ Damit verbunden ist die Ausblendung des ‚Nicht-Formalisierbaren‘ als eines essenziellen Teils der Wirklichkeit.

Einen Zugang mit hoher Erklärungskraft auf der Suche nach den nicht-digitalisierbaren Anteilen menschlichen Handelns und Denkens bietet das Konzept des erfahrungsgeleitet-subjektivierenden Handelns (vgl. Böhle 2009, 2017). Dieses Handlungskonzept unterscheidet systematisch zwischen einem *objektivierenden* (kognitiv-rationalen) Denken und Handeln und dem erfahrungsgeleitet-*subjektivierenden* Handeln (vgl. Abbildung 2). Im Zentrum steht dabei die Unterscheidung zwischen einem *expliziten* bzw. ‚objektivierbaren‘ Wissen und einem *impliziten Wissen* (tacit knowing) (vgl. Polanyi 1985), das seinen Gegenstand ‚subjektiviert‘ – und zwischen den jeweils resultierenden Handlungsformen.

Um *erfahrungsgeleitet-subjektivierendes Handeln* zu erfassen, wird bei der ‚Art der Vorgehens‘, beim ‚Gebrauch von Körper und Sinnen‘, bei der ‚Art des Denkens/Wissens‘ und bei der ‚Beziehung‘ zum Arbeitsgegenstand (Menschen und Dingen) jeweils zwischen einem rational kalkulierenden ‚objektivierenden‘ und einem ‚subjektivierenden‘ Zugang unterschieden. Die menschliche Fähigkeit, das eigene Handeln situativ und flexibel ohne langes Nachdenken und Planen an variierende Gegebenheiten anzupassen, geht im Wesentlichen auf implizites Wissen zurück.⁷ Wenn bei hohem Zeitdruck und großer Ungewissheit

6 Ähnlich der ‚Verschriftlichung‘ als zentrale Methode der Bürokratisierung.

7 Ein solches Handeln wird auch als ‚situierendes‘ Handeln (vgl. Suchman 2007) bezeichnet.

mit sogenannten ‚Feuerwehraktionen‘ gehandelt werden muss, ist diese besondere ‚Könnerschaft‘ (vgl. Neuweg 2015) entscheidend. Erfahrene Manager/-innen wie auch Fachkräfte führen diese oft auf Intuitionen, Emotionen und ‚Bauchgefühle‘ (vgl. Gigerenzer 2007) zurück. Diese ‚praktische Intelligenz‘ ist auch notwendig, um schnell das Verhalten anderer zu erkennen und flexibel darauf zu reagieren (vgl. Alkemeyer 2009).

Abbildung 2: Objektivierendes und subjektivierendes Handeln und Wissen



Quelle: eigene Darstellung

Das subjektivierende Arbeitshandeln wurde zunächst im Zuge von Untersuchungen des Arbeitshandelns in der chemischen Prozessindustrie erfasst. Im Umgang mit den großen Anlagen haben erfahrene Mitarbeiter/-innen sehr eindringlich geschildert, dass sie deren Zustand an der Spannung in der Luft erspüren, an aufgewühlten Farbnuancen in den Becken oder bereits bei der Anfahrt am Rauch erkennen und dass Gerüche wie Geräusche zusätzlich zu den Anzeigen von hoher Bedeutung für die Arbeit sind (vgl. Bauer et al. 2006, Böhle/Milkau 1988). An diese Forschungen hat mittlerweile eine große Zahl von empirischen Untersuchungen angeschlossen, die dieses Denken und Handeln in den verschiedensten Arbeitsbereichen (Wissensarbeit wie z.B. IT, Dienst-

leistungen wie z.B. Pflege usw.) gefunden haben und dessen Relevanz aufzeigen konnten.

An dieser Stelle soll festgehalten werden, dass sich das *objektiv-rationale Handeln* potenziell als direkt anschlussfähig an das ‚Maschinenhandeln‘ und die ‚Zeichenlogik‘ der Software erweist, während sich das *erfahrungsgeleitet-subjektivierende Handeln* der Explizierung seiner Wissensbasis systematisch verweigert – und sich damit auch dem direkten Zugriff der Digitalisierung entzieht. Denn durch das erfahrungsgeleitet-subjektivierende Handeln wird ein besonderes Erfahrungswissen aufgegriffen, das unmittelbar in praktisches Handeln eingebunden ist und hiervon nicht abgelöst kommuniziert, erworben und angewandt werden kann (vgl. Böhle 2015). Es lässt sich nicht ohne elementare Verluste in die ‚Zeichensprache‘ übersetzen und ist somit weder explizit verfügbar noch in üblicher Form als Wissen darstellbar und erkennbar. Dennoch kann subjektivierendes Handeln technisch unterstützt werden und Technik kann bei entsprechender Schnittstellengestaltung mit subjektivierendem Handeln in Interaktion treten. Entsprechend stehen hier inkorporierte Wissensbündel in Form eines grundlegenden und auf andere Kontexte übertragbaren Erfahrungswissens (und nicht bloße Erfahrung) im Mittelpunkt bzw. Kompetenzen und nicht allein Qualifikationen oder Fachwissen (vgl. Porschen-Hueck/Huchler 2016).

5.4 Zwischenfazit: Unhintergehbare, aber übergehbare Grenzen

Die Grenzen der Digitalisierung, denen (teil-)autonome Steuerung und adaptive lernende Systeme unterliegen, lassen sich also unterscheiden in: 1) Dilemmata und Nebenfolgen der Digitalisierung, die immer neue nicht-digitalisierte Aufgaben und Arbeiten entstehen lassen; 2) die Komplexität der physischen und sozialen Umwelt, die sich nicht vollständig technisch bearbeiten lässt; und 3) die Nicht-Formalisierbarkeit impliziten Wissens und Handelns, das sich nicht vollständig explizie-

ren und damit zugänglich für eine technisch-formale Bearbeitung machen lässt.

Dies markiert einerseits Grenzen, die der Digitalisierung realer Gegebenheiten und damit auch der Zugänglichkeit für adaptive bzw. lernende Maschinen auferlegt sind. Andererseits bedeutet dies nicht, dass diese Grenzen im digitalen Wandel auch tatsächlich Berücksichtigung finden. Vielmehr werden sie häufig ignoriert bzw. übergangen – mit nicht zu vernachlässigenden Folgen.

6. AUSBLENDUNG DER GRENZEN UND DIE FOLGEN

Die Ausblendung der Grenzen der Digitalisierung und Adaptivität geht mit typischen Fehlschlüssen einher, von denen im Folgenden ein paar wesentliche herausgegriffen werden.

6.1 Ersatz versus funktionales Äquivalent

Wenn vormals anders, z.B. durch menschliches Handeln, vollzogene Prozesse (teil-)automatisiert werden, werden sie nicht eins zu eins technisch reproduziert, sondern (allenfalls) möglichst detailgetreu simuliert. In der Regel beschränkt sich die technische Abbildung auf die Erstellung eines *funktionalen Äquivalents*.

Die Nicht-Beachtung dieses Unterschieds inszeniert einen vermeintlichen Wettbewerb zwischen Mensch und Technik anstelle einer Komplementarität. Kaum ein anderes Beispiel steht hierfür so sehr in der öffentlichen Diskussion, wie das Go-Spiel (siehe den Beitrag von Peter Brödner in diesem Band: S. 78ff.). Go gilt (im Vergleich zum Schach) aufgrund der enormen Komplexität als intuitives Spiel (immer noch in einem sehr eng kontrollierten bzw. potenziell abgeschlossenen Setting). Um die Komplexität zu beherrschen, greift nun KI bzw. ein

lernendes Programm auf statistische/mathematische Methoden und sehr aufwendige parallel laufende Prozesse zurück.⁸ Die Resultate werden auf Basis neuer Methoden, einer stark gestiegenen Rechenleistung und vorhandener Daten erzielt. Menschliche Spieler/-innen könnten auf diese Weise nicht agieren, sondern sind auf Intuition angewiesen. Intuition wird hier nicht einmal technisch simuliert, sondern nur funktional ersetzt. Vielmehr agiert ‚die KI‘ auf eine andere Weise, bei der die Technik (z.B. der Taschenrechner) dem Menschen schon lange überlegen ist: die Durchführung formaler Rechenoperationen. Ein weiteres Beispiel sind ‚empathische‘ Systeme wie Sprach- und Gesichtserkennungsprogramme, die auf Gefühle reagieren. Sie simulieren nicht Empathie, sondern ersetzen diese Kompetenz durch mathematische Modelle bzw. Korrelationen. Auch semantische Technologien oder Bilderkennungssysteme erfassen nicht selbst die Bedeutung von Wörtern und Bildern, sondern übersetzen sie in Zahlen, um sie dann mit vorhandenen Daten, Verknüpfungen und wahrscheinlichen Kombinationen abzugleichen. Damit können sie sich wahrscheinlichen Bedeutungen annähern und sogar zu neuen Werken kombinieren, ohne diese selbst zu erfassen oder gar zu ‚verstehen‘; inklusive der eigenen vermeintlichen ‚Kreativität‘.

Auch lernen Mensch und Maschine sehr unterschiedlich.⁹ Ohne auf diesen zentralen Aspekt genauer eingehen zu können, fällt auf, dass

-
- 8 Mittels extrem vieler (40×10^3) parallel durchgeführter Zufallspartien werden für bestimmte Stellungen Zugkombinationen mit hoher Erfolgswahrscheinlichkeit identifiziert. Zudem greift das System auf Metadaten zurück, die durch zwei gegenseitig trainierte lernende Systeme ständig erweitert werden, usw.
- 9 Als ein Beispiel für das Erlernen von motorischen Tätigkeiten lässt sich das Fahrradfahren nehmen: Der technisierbare (objektivierende, auf explizitem Wissen beruhende) Zugang zum Fahrradfahren wäre, die eigene Lage im Raum ständig zu überprüfen, den nächsten Schritt zu planen und so in Sekundenbruchteilen die Balance durch Lenkerbewegungen, Körperlage und Treten aufrechtzuerhalten. Diese Handlungsabfolgen könnten dann aufge-

Menschen sich im ersten Schritt wesentlich flexibler und schneller eine neue Aufgabe adaptieren können, während technische lernende Systeme sehr viele Versuche und den Aufbau einer extrem großen Datenmenge benötigen. Dann aber sind technische Systeme besonders gut geeignet, einen ‚trainierten‘ Schritt immer gleich und immer schneller zu vollziehen. Auch können technische Systeme ihr ‚explizites Wissen‘ bzw. ihre angesammelten Daten einfach auf ein anderes System übertragen und so unschlagbar schnell und genau voneinander lernen. Eine Herausforderung ist es z.B., einen Zylinder in eine enge Form zu stecken: Der Mensch trifft schnell das Loch und ‚rüttelt‘ den Zylinder rein, damit er sich nicht verklemmt. Ein Roboterarm braucht eine ‚naïve‘ Annäherung an den optimalen Prozess. Hat er diesen jedoch einmal eingestellt bzw. erlernt, kann er sehr schnell mit nicht mehr sichtbaren Vibrationen in ständig gleicher Wiederholung ausgeführt und auf andere Roboterarme übertragen werden. Der Roboterarm ‚überholt‘ an dieser Stelle die haptischen Fähigkeiten des Menschen, da mit der ständigen Wiederholung die Anforderung von Adaptivität weg hin zu Reproduzierbarkeit geht.

Die Möglichkeit einer technischen Abbildung allen Denkens und Handelns wird derzeit gern als eine bloße Frage der Zeit behandelt, unter Heranziehung zahlreicher Beispiele für vermeintlich kreative oder empathische ‚Künstliche Intelligenz‘. Dabei bleibt meist unthematisiert, dass zwar scheinbar das Gleiche getan wird, jedoch auf entscheidend andere Art und Weise. Manuelle Tätigkeiten, Wissensarbeit wie

zeichnet, in einem enormen Datensatz gebündelt und in einen Algorithmus überführt werden, anhand dessen andere Systeme Fahrradfahren lernen und optimieren können. Menschen lernen das Fahrradfahren anders: Draufsetzen, anschieben, relativ schnell alleine fahren. Auch während des Fahrens wäre es fatal, erst über die momentan notwendige ausgleichende Abstimmung zwischen Gerät und Körper nachzudenken und dann zu handeln. Vielmehr müssen Handeln und Denken ineinanderfließen. Unbewusst ist dieses Handeln deshalb nicht. Es folgt nur nicht der klassischen Abfolge bewusst kalkuliert getroffener Entscheidungen: erst Denken, dann Handeln.

auch personenbezogene Interaktionsprozesse werden höchstens technisch simuliert, in der Regel jedoch nur funktional ersetzt. So oder so eignet sich die entsprechende Technologie weder die mit der Tätigkeit verbundenen menschlichen Kompetenzen an noch ist der Weg, *wie* die Tätigkeit vollzogen wird, der gleiche. Das unterschiedliche Vorgehen ist verbunden mit unterschiedlichen Voraussetzungen, was die Prozessbedingungen und die Prozessgestaltung angeht, und mit entsprechend unterschiedlichen Folgewirkungen. So müssen die technischen Voraussetzungen zur Datenerfassung (‚Wahrnehmung‘) und -verarbeitung erst geschaffen werden, die Datenmenge, die benötigt wird, ist eine andere, und auch die Art der Verarbeitung der Daten unterscheidet sich grundlegend.

Diese Unterschiede können manchmal in der Praxis vernachlässigbar sein, oftmals erweisen sie sich jedoch als wesentlich für die Qualität des Resultats, für die Flexibilität bzw. Stabilität des Prozesses, für die einhergehenden Nebenfolgen für Arbeit und Gesellschaft. Was fällt mit der (Teil-)Automatisierung weg bzw. bleibt liegen? Was kommt neu hinzu? Entscheidend ist hier die limitierte Explizierbarkeit bzw. Formalisierbarkeit realer Gegebenheiten.

Es kann konstatiert werden, dass (teil-)autonome Systeme zwar zum Teil annähernd dasselbe tun, aber auf eine andere Art und Weise. In der Folge muss es darum gehen, genau zu betrachten, was dies im Resultat bedeutet und welche Aufgaben am besten *wie* und damit auch *durch was/wen* erledigt werden sollten. Es braucht also Kriterien für eine neu begründete *Arbeitsteilung* zwischen lernendem System und humaner Arbeitskraft. Diese können entlang der Grenzen der Digitalisierung bzw. Adaptivität technischer Systeme entwickelt werden.

6.2 Überschätzung des Substituierungspotenzials durch Ausblendung des Informellen

Fragen nach der Ersetzbarkeit menschlicher Arbeit durch digitale Technologien und den damit verbundenen Folgen für Arbeit und Gesellschaft stehen im Zentrum des öffentlichen Digitalisierungsdiskur-

ses. Die Einschätzungen zum Substitutionspotenzial variieren stark. Jedoch scheinen sie sich zunehmend auf ein Szenario einzuspielen, das nicht von einem ersatzlosen Wegfall von Arbeit ausgeht, sondern von einer Verschiebung von Tätigkeitsfeldern (vgl. IAB-Forschungsbericht 2016). Eine Phase der Überhöhung und teilweise auch Instrumentalisierung des Diskurses scheint einem differenzierteren Hinsehen zu weichen; einhergehend mit einem sinkenden geschätzten Substitutionspotenzial.¹⁰

Nach wie vor stehen, wenn über besonders gefährdete Tätigkeiten gesprochen wird, sogenannte Routinetätigkeiten (in Produktion, Sachbearbeitung und Dienstleistung) im Fokus. Dabei wird allerdings die Komplexität auch dieser Tätigkeiten systematisch unter- und entsprechend das Substitutionspotenzial überschätzt (vgl. Pfeiffer/Suphan 2015). Auch hier reicht der Blick auf die formal erfassbaren Arbeitsanteile nicht aus. Wenn vermeintliche Routinetätigkeiten (teil-)automatisiert werden, werden die informellen Anteile der Arbeit, die stillen Leistungen, die notwendigen zusätzlichen Kompetenzen und das Erfahrungswissen oft erst wirklich offensichtlich, z.B. indem angrenzende Arbeitsbereiche die nicht erfasste, aber dennoch anfallende Arbeit auffangen müssen. In der Regel steigt dann die Arbeitsbelastung im Umfeld an. Zugleich macht es wenig Sinn, aus einer Arbeit alle Routineanteile zu entfernen. Sie würde damit extrem intensiviert, da Routine auch entlastend wirkt. Hinzu kommt, dass Routine oftmals auch erst durch angesammelte Kompetenzen entwickelt und wirksam wird.

Insofern ist es fruchtbarer, auf der Suche nach Substitutionspotenzialen nicht von Routineanteilen zu sprechen, sondern von den formalisierbaren (vs. nicht-formalisierbaren) Anteilen von Arbeit; inklusive derer, die durch Metadaten abbildbar sind. Kreativität, Spieltrieb, In-

10 Z.B. von 50 Prozent (USA) bei Carl B. Frey und Michael A. Osborne (2013) auf 23 Prozent (für die USA) bei McKinsey Global Institut (2017) oder 9 Prozent für alle OECD-Staaten im Schnitt bei Melanie Arntz, Terry Gregory und Ulrich Zierahn (2016) unter Betrachtung von Teiltätigkeiten und Berücksichtigung normaler Anpassungsprozesse.

novation, aber auch Empathie, Authentizität, das Aushalten von und Handeln in Widersprüchen gelten als Eigenschaften, die sich nicht wirklich formal ‚steuern‘, sondern nur durch passende Rahmenbedingungen fördern lassen. Entsprechende Tätigkeiten und Berufe bergen deshalb auch ein erhöhtes Verhandlungspotenzial gegenüber Arbeit- oder Auftraggeber/-innen. Diese typisch menschlichen Kompetenzfelder stellen eine besondere Herausforderung für die Entwicklung (teil-)autonomer Systeme dar, die aktuell insbesondere mittels lernender Systeme aufgegriffen wird (s.o.).

Sie sind jedoch nur Ausschnitte des subjektivierenden Arbeitshandelns, das sich insgesamt der Formalisierung verwehrt und eher das *Wie* der Arbeit als das inhaltliche *Was* beschreibt. Es kann sich potenziell in jeder Arbeit äußern, auch in Routinetätigkeiten und vor allem in der Facharbeit und eben nicht nur in ‚kreativen‘ oder personenbezogenen Berufen, wo Empathie gefragt ist. Dieses *Wie* setzt dem Herausrücken des Menschen aus den Arbeitsprozessen und der Idee des Entscheidungstreffens auf Basis von lernenden Systemen vorbereiteter Entscheidungsbündel gewisse Grenzen, da spezifische Handlungsformen und Kompetenzen nicht mehr direkt im Arbeitsprozess eingebracht werden. Eine Alternative ist daher eine komplementäre Mensch-Technik-Kollaboration im Arbeitsprozess, was neue Anforderungen an die Technikgestaltung stellt (s.u.).

Erneut kann gefolgert werden: Es existieren Grenzen der (Teil-)Automatisierung in jeder Tätigkeit. Dennoch können die Tätigkeiten von technischen Systemen übernommen werden, wobei immer etwas übrig bleibt bzw. wegfällt und Neues hinzukommt. Denn mit diesem Schritt verändert sich die Art und Weise der Bearbeitung, mit entsprechenden (mehr oder weniger großen) Folgen.

6.3 Anpassung der Umwelt an die Bedarfe des technischen Systems

Eine typische Reaktion auf wahrgenommene Grenzen der technischen Beherrschbarkeit bzw. der Adaptionsfähigkeit ist die Standardisierung der Umgebung mit dem Ziel, die Umwelt für formale Systeme anschlussfähig und beherrschbar zu machen. Dies kann auf Kosten der Vielfalt, der Qualität und der individuellen Freiheiten gehen. So lässt sich der autonome Straßenverkehr in einer hoch reglementierten Umgebung am einfachsten umsetzen; z.B. wenn sich nur mit entsprechender Technologie ausgestattete Fahrzeuge auf der Straße befinden, die Systeme alle miteinander kompatibel sind, Straßen nur an Ampeln überquert oder gekreuzt werden können etc. Auch sind die Nutzer/-innen die besten Kund/-innen, die den antizipierten und geplanten Nutzerprofilen entsprechen. Entsprechend gilt es, das Nutzerverhalten (inkl. Nutzerbedürfnisse) nicht nur schon bereits im Vorfeld zu erfassen, sondern auch zu steuern (vgl. Lobe 2018).

Eine nicht gering zu schätzende Gefahr besteht darin, dass sich Nutzer/-innen zu sehr an die Bedarfe und Unzulänglichkeiten des Systems (bzw. der Hersteller) anpassen; z.B. wenn emotionssensible IT auf Basis von Sprach- und Gesichtserkennung für Einstellungstests verwendet wird, den Lernerfolg im Unterricht bemisst oder individuelle Kundenwünsche beim Einsteigen in das autonome Taxi vom Gesicht abliest.¹¹ KI-Systeme könnten dann Menschen auf gute erkennbare oder positiv sanktionierte Sprache und Mimik konditionieren.¹² Berater/-innen würden Menschen für wichtige Einstellungen auf die jewei-

11 Bei allen hier aufgeführten Beispielen handelt es sich um aktuelle Testprojekte.

12 Als Beispiel kann der boomende Markt an digitalen HR-Analyse-Tools im Personalmanagement genannt werden, z.B. Persönlichkeitsprofilierung auf Basis von Schriftstücken, Mimik und Gestik oder eines beliebigen gesprochenen Textes. Sie basieren auf Typisierungen und sind – einmal in ihrer Logik durchschaut – auch manipulationsanfällig.

ligen KI-Systeme trainieren, Schüler/-innen würden Aufmerksamkeit fingieren und damit das Taxi die Türe verriegelt oder schneller fährt, würden sich übertriebene Gestiken zur Steuerung etablieren. Die Vielzahl von Formen, wie Menschen Emotionen (z.B. Freude) ausdrücken können, steht einer limitierten Anzahl von Kategorien bzw. ‚Schubladen‘, mit denen lernende Systeme auf Basis von Wahrscheinlichkeiten arbeiten, gegenüber. Dies birgt typische Gefahren der Fremd- und Selbst-Disziplinierung bzw. normierenden Anpassung an die Bedarfe und Limitierungen technischer Systeme.

6.4 Akzeptanz anstelle echter Partizipation

Nicht nur eine überzogene bzw. entfremdende Anpassung des Menschen an die Bedarfe der Technik, sondern auch der umgekehrte Prozess birgt Gefahren, insbesondere wenn Anpassung mit Nachahmung vertauscht wird und die Grenzen der technischen Adaptivität und die Unterschiede zwischen menschlichem Handeln und dessen digitaler Simulation unberücksichtigt bleiben.

Seit einiger Zeit wird intensiv über die Agenda ‚Humanizing Technology‘ diskutiert. Dabei wird argumentiert, dass Technikentwicklung stärker von dem Kunden/der Kundin bzw. ‚von unten‘ (z.B. von dem Mitarbeiter/der Mitarbeiterin) aus gedacht werden muss. So soll unter anderem darauf reagiert werden, dass Technik oftmals ‚de-humanisierende‘ bzw. entfremdende Wirkungen erzeugt, wenn sie Nutzer/-innen zwingt, sich ihrer Logik und der ihrer Entwickler/-innen zu unterwerfen. Daraus wird häufig gefolgert, dass Technik dem Menschen ähnlicher gestaltet werden müsse, damit die Mensch-Maschine-Interaktion (MMI) weniger entfremdend wirkt. Dies ist jedoch zu hinterfragen. Natürlich macht es Sinn, Sprachsteuerungen und andere Schnittstellen in der MMI zugänglicher, niedragschwelliger und zum Teil auch intuitiver zu gestalten. Aber die Losung, dass *Ähnlichkeit* Entfremdung bzw. De-Humanisierung entgegenwirkt und Vertrauen in der Interaktion erzeugt, läuft in ihrer Konsequenz fehl – da sie weiterhin an der Substitution des Menschen festhält und hierdurch die Tech-

nik in den Mittelpunkt stellt. Diese Perspektive hält an der Agenda der Kolonialisierung der Lebenswelt durch Technik fest, statt eine echte Unterstützungs- und vor allem Entwicklungsperspektive menschlicher Fähigkeiten zu eröffnen und über diesen Weg neue technische Entwicklungspfade zu erschließen.¹³ Beides kann nur gelingen durch eine abgestimmte *Arbeitsteilung* und *wechselseitige Lernförderlichkeit* im sozio-technischen System. Basis hierfür ist eben nicht Ähnlichkeit. Auch Akzeptanz und Erfolg der Interaktion und Kooperation zwischen technischem System und Menschen lassen sich durch eine echte Befähigung und Ermächtigung der Nutzer/-innen und über erkennbare Nützlichkeit, Anwendbarkeit und (bedingte) Kontrollierbarkeit der Effekte eher erreichen als durch eine – notwendigerweise eingeschränkte und damit nie authentische – Ähnlichkeit zwischen technischem System und Mensch.¹⁴

13 Nicht nur die Entwicklungskonzepte der ‚Customer Experience‘ und des ‚Human-Centric Product Design‘ (z.B. Apple; vgl. Hiner 2011), sondern auch die der ‚Usability‘ oder ‚intuitiven Nutzerführung‘ allgemein sind noch weit entfernt von einer echten Aneignungs- und Empowermentperspektive. So bleiben weite Teile menschlicher Wahrnehmung und Verarbeitung kaum adressiert und werden nicht gezielt gefördert. Selbst Don Norman, der als Designer bei Apple den Begriff ‚User Experience‘ prägte, sieht noch erheblichen Handlungsbedarf und identifiziert aktuelle Fehlentwicklungen, z.B. da entdeckendes Lernen durch intuitive Nutzerführung nicht mehr möglich sei (vgl. Heuer 2013).

14 Statt z.B. in der Produktion Leichtbauroboterarme mit sieben Gelenken künstlich auf die Bewegungen des menschlichen Arms zu beschränken (um Vertrauen zu erzeugen und Unfälle zu vermeiden), könnte es die Handhabung des ‚Werkzeugs‘ und die Bindung zu ihm verbessern, wenn der ‚Nutzer‘/die ‚Nutzerin‘ sich gerade die besonderen ergänzenden Möglichkeiten der Technik aneignen kann, um Dinge zu tun, die er/sie nicht von selbst kann. Die teils starke (auch emotionale) Bindung zum privaten Kfz, zum großen Bagger, zum gut in der Hand liegenden Hammer (als ‚verlängerter Arm‘) wird eben dadurch hergestellt, dass man durch das Objekt zu an-

Zudem lassen sich die Aktivitäten eines unabhängig interagierenden Systems (z.B. Leichtbaurobster) wesentlich besser nachvollziehen, wenn sie offen der rationalen Logik expliziten Wissens folgen, also z.B. transparenten Regeln, wenn sie also klar vorhersagbar sind und keine ‚menschliche‘ Charaktermaske tragen, die die Interaktion verkompliziert.¹⁵ Ähnlichkeit kann sogar technisch dysfunktional sein und unnötige Risiken eröffnen, etwa wenn für Flugzeugcockpits humanoide Roboter entwickelt und so unnötig zusätzliche Schnittstellen aufgemacht werden, statt sich auf das Fluggerät als eigentliches (teil-)autonomes System zu konzentrieren.

Daraus lässt sich folgern: Wenn (teil-)autonome lernende Systeme umfassend adaptiv sein sollen, ist es notwendig, zunächst die Beschränkungen der Adaptivität zu berücksichtigen und dann die Systeme dort für anderweitige Lösungen zu öffnen, insbesondere für menschliches Handeln. Eine durch das technische System (bzw. dessen Programmierung) vorgesehene Auswahl an vorbereiteten Entscheidungsmöglichkeiten ist nicht ausreichend, da so nur in sehr begrenztem Maße Komplexität in die Situation genommen wird. ‚Offene Schnittstellen‘, die nicht auf die durchgängige Beherrschung bzw. technische Koordination des Nutzungs- oder Interaktionsprozesses abzielen, müssen das Steuer bzw. die ‚Handlungsträgerschaft‘ (siehe auch den Beitrag von Schulz-Schaeffer in diesem Band) gezielt abgeben und an nahezu beliebiger Stelle wieder aufgreifen können. Dies würde jedoch bedeuten, die Nutzer/-innen umfassend einzubinden – eben genau da, wo menschliches Handeln funktional (oder auch normativ) angezeigt ist. Dies stellt eine enorme Herausforderung für die Entwicklung entsprechender Systeme dar. Es ginge darum, die Defizite rein technischer Adaptivität zu erkennen, um diese dann durch Hinzunahme menschl-

sonsten unmöglichen Handlungen und zur Ausübung von Kompetenzen befähigt wird.

- 15 So kann es zielführender sein, wenn ein mobiles System in der Intralogistik immer entweder gleich ausweicht oder stehen bleibt, statt die Bewegungen des menschlichen Gegenübers zu ‚interpretieren‘ und vorherzusagen.

chen Handelns zu ergänzen – oder umgekehrt menschliches Handeln zu analysieren und dort zu ergänzen, wo die technische Bearbeitung als sinnvoll eingestuft wird. Insgesamt würde so die Adaptivität des sozio-technischen Systems steigen; und damit die Handlungsfähigkeit in unsicheren und komplexen Situationen.

6.5 Zwischenfazit: Assimilierende Adaptivität

Die aktuelle Diskussion um den digitalen Wandel von Arbeit, insbesondere angesichts (teil-)autonomer Steuerung und lernender Systeme, ist geprägt durch eine Anzahl wesentlicher Fehlschlüsse. So wird 1.) die Technisierung eines vormals anders (in der Regel durch menschliches Zutun) organisierten Prozesses als gleichwertiger Ersatz angenommen, aber nicht berücksichtigt, dass zum Erreichen des Effekts (,funktionales Äquivalent‘) ein anderer Weg gewählt wird. Technische Systeme sind nicht deshalb intuitiv, kreativ oder empathisch, weil sie für Aufgaben eingesetzt werden, die Menschen mittels Intuition, Kreativität oder Empathie bearbeiten. 2.) Das Substitutionspotenzial von menschlichem Arbeitshandeln wird systematisch dadurch überschätzt, dass die informellen Leistungsanteile und impliziten Kompetenzen ausgeblendet bleiben und nur die explizierbaren Wissensanteile bzw. die formalisierbaren Arbeitsanteile betrachtet werden, da sich auch nur diese digitalisieren lassen. 3.) Darüber hinaus besteht die Tendenz, die Umwelt an die Bedarfe der technischen Systeme anzupassen, um die Komplexität für die notwendigerweise Eingeschränktheit adaptiver Systeme bearbeitbar zu gestalten. 4.) Diese Tendenz wird noch zusätzlich dadurch verstärkt, dass unter dem Label ,Humanizing Technology‘ die Technik dem Menschen ähnlich gestaltet wird, anstatt systematisch auf eine wechselseitige Ergänzungsperspektive zu setzen. Ein Weg zu Letzterem wären ,*offene adaptive sozio-technische Systeme*, die die Situationskoordination bzw. Handlungsträgerschaft situativ abgeben und wieder aufgreifen können, jenseits von eng geschnittenen Optionen bzw. Wenn-dann-Regeln (s.u.).

Angesichts dieser aktuellen Tendenzen kann die deutliche Dominanz einer Beherrschbarkeitsideologie in der Technikentwicklung diagnostiziert werden, die mit einer Überschätzung von in Aussicht gestellten Skalierungseffekten durch digitale Systeme einhergeht. Während die technische Adaptivität (teil-)autonomer, insbesondere lernender Systeme steigt, bleiben die Grenzen der Digitalisierung oftmals ausgeblendet. Insofern kann von einer ‚*assimilierenden Adaptivität*‘ gesprochen werden, bei der Umweltbedingungen soweit möglich in die Eigenlogik des Systems übersetzt werden und Nicht-Übersetztes (und insbesondere Nicht-Übersetzbares) außen vor bleibt. In der Folge kommt es zu typischen Fehlschlüssen von technischer Machbarkeit im Labor, in der Simulation, in der Einzelumsetzung auf die Umsetzung in der Breite, insbesondere durch die Ausblendung der Komplexität der Praxis und der Reichhaltigkeit menschlichen Arbeitshandelns.

7. FAZIT: KOMPLEMENTÄRE ADAPTIVITÄT DURCH OFFENE SYSTEME UND HUMANE ARBEIT

Zusammenfassend lässt sich sagen, dass angesichts der Grenzen der Digitalisierung vordringlich die Steigerung einer *assimilierenden Adaptivität* beobachtbar ist, bei der (teil-)autonome Systeme befähigt werden, sich ihrer Umwelt dadurch anzupassen, dass sie diese in die formale Eigenlogik des Systems überführen. Dies steht dem Erreichen einer *komplementären Adaptivität*, für die ein wechselseitiger Kooperationszusammenhang zwischen technischem System und anderen ‚Systemen‘ in seinen Umwelten wesentlich ist, konflikthaft gegenüber. Letztere würde jedoch eine *Steigerung* der Adaptivität technischer (insbesondere lernender) Systeme ermöglichen. Als Resultat könnten ‚*offene*‘ *sozio-technische Systeme* entstehen, in welchen die Situationskoordination bzw. Handlungsträgerschaft situativ je nach Bedarf und Kompetenzen der ‚Akteure‘ zwischen Mensch und Technik übertragen wird. Hierfür müssten (teil-)autonome Systeme in der Lage sein, die

Koordination des Prozesses tatsächlich abzugeben und wieder aufzugreifen, und zwar ohne vorher die möglichen Anschlussstellen festzulegen.¹⁶ Damit würden sehr hohe Anforderungen an die Technikgestaltung in Richtung Adaptivität einhergehen, z.B. eine permanente Umprogrammierung im Interaktions-/Nutzungsprozess. Lernende Systeme könnten hier neue Wege ermöglichen.

Dies betrifft nicht nur die Gestaltung der Mensch-Maschine-Interaktion, sondern die gesamte Arbeitsorganisation bzw. das gesamte betriebliche Setting, in das die Technologie wie auch das Arbeitshandeln eingebunden sind (vgl. Ittermann et al. 2016). Denn die Einbettung technischer Systeme in den gesamten Anwendungskontext ist entscheidend für die Ausformung der Interaktion und deren Folgen.

Grundlage für beides, die technische und organisationale Gestaltung, ist eine (erneute) Auseinandersetzung mit den Grenzen der Digitalisierung auf Basis aktueller Technologien und aus dieser Perspektive eine Neubetrachtung der Arbeitsteilung zwischen Mensch und Technik. Diese Auseinandersetzung ist dringend nötig angesichts dessen, dass die aktuellere Diskussion zum digitalen Wandel (Industrie 4.0, lernende Systeme etc.) zwar geprägt ist durch Statements, dass der Mensch ‚weiterhin‘ „im Mittelpunkt“ der Arbeit stehen soll (z.B. Promotorengruppe Kommunikation der Forschungsunion Wirtschaft – Wissenschaft 2013: 24; Spath et al. 2013: 129f.; vgl. auch acatech 2016; BMAS 2016; Plattform Industrie 4.0 2015), diese *normative* Setzung aber kaum bis gar nicht theoretisch begründet oder zumindest *funktional* untermauert wird. Zwar distanziert man sich von dem gescheiterten Programm Computer Integrated Manufacturing (CIM) der 1980er Jahre und dem dort propagierten Leitbild der ‚mensenleeren Fabrik‘. Gleichzeitig wird aber argumentiert, dass jetzt die technische

16 In einem digitalen technischen System müssen die Anschlussstellen immer in der Systemarchitektur vorgesehen sein oder durch sie zumindest ermöglicht werden. Insofern sind sie immer formal festgelegt. Die Herausforderung ist nun, auch unvorhergesehene Ausformungen (Wege, Inhalte, Zeitpunkte etc.) der Interaktionen mit dem System zu ermöglichen.

Basis dafür bestehe, die Visionen von CIM schrittweise umzusetzen (exemplarisch Gausemeier 2015; Scheer 2012; Stich 2016). Mit dezentralen lernenden Systemen, so heißt es, ließen sich die CIM-Ideen doch noch realisieren, nun aber mit Blick auf ganze Wertschöpfungssysteme. Das Ideal ist die Reduktion der Transaktionskosten auf null. Dabei werden die Besonderheiten menschlichen Arbeitshandelns (inklusive des damit verbundenen Erfahrungswissens) nicht nur auf dem Shopfloor, sondern letztlich in allen Bereichen durch ‚Machine Learning‘ bzw. lernende Systeme als ersetzbar angesehen.

Ziel des Beitrags ist es, einen rein technikfokussierten Blick zu erweitern auf den sozio-technischen Zusammenhang. Auf Basis einer ersten Systematisierung von Grenzen der Digitalisierung wurden ein assimilierendes und ein komplementäres Adaptivitätsverständnis unterschieden und einander gegenübergestellt. Die Logik, auf erhöhte Unsicherheit und Komplexität mit Dezentralisierung bzw. einer Steuerung aus dem Prozess heraus zu reagieren, um Adaptivität und damit Flexibilität zu steigern, liegt nahe und ist auch nicht neu. Sie kennzeichnet den aktuellen Wandel von Arbeit insgesamt. Ein weiterer Schritt wäre die Steigerung der Adaptivität (teil-)autonom lernender Systeme in Richtung offener Systeme und komplementärer Adaptivität. Dies würde den Fokus auf die Gesamt-Adaptivität des sozio-technischen Systems legen und könnte ein Schlüssel sein für eine humanorientierte Technikgestaltung.

LITERATUR

- acatech (2016): Kompetenzen für Industrie 4.0. Qualifizierungsbedarfe und Lösungsansätze (= acatech POSITION), München: Herbert Utz.
- Alkemeyer, Thomas (2009): „Handeln unter Unsicherheit – vom Sport aus beobachtet“, in: Fritz Böhle/Margit Weihrich (Hg.): Handeln unter Unsicherheit, Wiesbaden: VS Verlag, S. 183-202.

- Arntz, Melanie/Gregory, Terry/Zierahn, Ulrich (2016): The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis (ZEW Mannheim, Universität Heidelberg), https://www.oecd-ilibrary.org/social-issues-migration-health/the-risk-of-automation-for-jobs-in-oecd-countries_5j1z9h56dvq7-en vom 20.8.2018.
- Bainbridge, Lisanne (1983): „Ironies of automation“, in: *Automatica* 19/6, S. 775-779.
- Bauer, Hans G./Böhle, Fritz/Munz, Claudia/Pfeiffer, Sabine/Woicke, Peter (2006): *Hightech-Gespür. Erfahrungsgeleitetes Arbeiten und Lernen in hoch technisierten Arbeitsbereichen*, Bielefeld: Bertelsmann.
- Baukrowitz, Andrea/Berker, Thomas/Boes, Andreas/Pfeiffer, Sabine/Schmiede, Rudi/Will, Mascha (Hg.) (2006): *Informatisierung der Arbeit – Gesellschaft im Umbruch*, Berlin: edition sigma.
- Beck, Ulrich (1986): *Risikogesellschaft. Auf dem Weg in eine andere Moderne*, Frankfurt a.M.: Suhrkamp.
- Birken, Thomas/Pelizäus-Hoffmeister, Helga/Schweiger, Petra/Sontheimer, Rainer (2018): „Technik für ein selbstbestimmtes Leben im Alter – eine Forschungsstrategie zur kontextintegrierenden und praxiszentrierten Bedarfsanalyse“ [96 Absätze], in: *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 19/1, Artikel 3, <http://dx.doi.org/10.17169/fqs-19.1.2871> vom 20.8.2018.
- Böhle, Fritz (2009): „Weder rationale Reflexion noch präreflexive Praktik. Erfahrungsgeleitet-subjektivierendes Handeln“, in: Fritz Böhle/Margit Wehrich (Hg.): *Handeln unter Unsicherheit*, Wiesbaden: VS Verlag, S. 203-230.
- Böhle, Fritz (2015): „Entscheidungs- und Handlungsfähigkeit mit Ungewissheit“, in: *praeview. Zeitschrift für innovative Arbeitsgestaltung und Prävention* 2/3, S. 9.
- Böhle, Fritz (Hg.) (2017): *Arbeit als Subjektivierendes Handeln. Handlungsfähigkeit bei Unwägbarkeiten und Ungewissheit*, Wiesbaden: VS Verlag.
- Böhle, Fritz/Huchler, Norbert (2016): „Cyber-Physical Systems and Human Action. A Re-Definition of Distributed Agency Between

- Humans and Technology, Using the Example of Explicit and Implicit Knowledge“, in: Houbing Song/Danda B. Rawat/Sabine Jeschke/Christian Brecher (Hg.), *Cyber-Physical Systems: Foundations, Principles, and Applications*, Amsterdam: Elsevier, S. 115-127.
- Böhle, Fritz/Milkau, Brigitte (1988): *Vom Handrad zum Bildschirm. Eine Untersuchung zur sinnlichen Erfahrung im Arbeitsprozeß*, Frankfurt a.M.: Campus.
- Böhle, Fritz/Pfeiffer, Sabine/Porschen, Stephanie/Sevsay-Tegethoff, Nese (2011): „Herrschaft durch Objektivierung. Zum Wandel von Herrschaft in Unternehmen“, in: Wolfgang Bonß/Christoph Lau (Hg.): *Herrschaft durch Uneindeutigkeit*, Weilerswist: Velbrück, S. 244-283.
- Böhle, Fritz/Porschen, Stephanie (2012): „Verwissenschaftlichung und Erfahrungswissen. Zur Entgrenzung, neuen Grenzziehungen und Grenzüberschreitungen gesellschaftlich anerkannten Wissens“, in: Ulrich Wengenroth (Hg.), *Grenzen des Wissens – Wissen um Grenzen*, Weilerswist: Velbrück, S. 154-192.
- Böhle, Fritz/Rose, Helmuth (1992): *Technik und Erfahrung. Arbeit in hochautomatisierten Systemen*, Frankfurt a.M.: Campus.
- Boes, Andreas (2005): „Informatisierung“, in: SOFI/IAB/ISF München/INIFES (Hg.): *Berichterstattung zur sozioökonomischen Entwicklung in Deutschland. Arbeits- und Lebensweisen. Erster Bericht*, Wiesbaden: VS Verlag, S. 211–244.
- Brödner, Peter (2015): „Industrie 4.0 und Big Data – wirklich ein neuer Technologieschub?“, in: Hartmut Hirsch-Kreinsen/Peter Ittermann/Jonathan Niehaus (Hg.), *Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen*, Baden-Baden: Nomos, S. 231-250.
- Bundesministerium für Arbeit und Soziales (BMAS) (Hg.) (2016): *Weissbuch Arbeiten 4.0*, Berlin: Bundesministerium für Arbeit und Soziales.
- Fink, Robin D./Weyer, Johannes (2014): „Interaction of Human Actors and non Human Agents. A Sociological Simulation Model of Hy-

- brid Systems“, in: Science, Technology & Innovation Studies 10/1, S. 47-64.
- Gausemeier, Jürgen (2015): „Computer Integrated Manufacturing (CIM)“, in: Norbert Gronau/Jörg Becker/Elmar J. Sinz/Leena Suhl/Jan M. Leimeister (Hg.): Enzyklopädie der Wirtschaftsinformatik. Online-Lexikon, [http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/informationssysteme/Sektorspezifische-Anwendungssysteme/Computer-Integrated-Manufacturing-\(CIM\)](http://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/informationssysteme/Sektorspezifische-Anwendungssysteme/Computer-Integrated-Manufacturing-(CIM)) vom 17.8.2017.
- Gigerenzer, Gerd (2007): Bauchentscheidungen. Die Intelligenz des Unterbewussten und die Macht der Intuition, München: Bertelsmann.
- Grote, Gudela (2015): „Gestaltungsansätze für das komplementäre Zusammenwirken von Mensch und Technik in Industrie 4.0“, in: Hartmut Hirsch-Kreinsen/Peter Ittermann/Jonathan Niehaus (Hg.): Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen, Baden-Baden: Nomos, S. 131-146.
- Heuer, Steffan (2013): „Das Thema Usability ist für mich erledigt“, in: brand eins 7, S. 71-74.
- Hiner, Jason (2011): Humanizing Technology: The 100-Year Legacy of Steve Jobs, ZDnet vom 11.10.2011, <http://www.zdnet.com/article/humanizing-technology-the-100-year-legacy-of-steve-jobs/> vom 5.11.2015.
- Hirsch-Kreinsen, Hartmut/Ittermann, Peter/Niehaus, Jonathan (Hg.) (2015): Digitalisierung industrieller Arbeit. Die Vision Industrie 4.0 und ihre sozialen Herausforderungen, Baden-Baden: Nomos.
- Huchler, Norbert (2016): „Re-Taylorisierung und Subjektivierung. Die Risiken digitaler Arbeit und Leitlinien für eine nachhaltige Arbeits- und Technikgestaltung“, in: DGUV FORUM. Fachzeitschrift für Prävention, Rehabilitation und Entschädigung der Deutschen Gesetzlichen Unfallversicherung (DGUV) 11, S. 12-16.
- Huchler, Norbert (2018): „Die Grenzen der Digitalisierung. Begründung einer Arbeitsteilung zwischen Mensch und Technik und Im-

- plikationen für eine humane Technikgestaltung“, in: Josephine Hofmann (Hg.): Arbeit 4.0 – Digitalisierung, IT und Arbeit, Wiesbaden: Springer Vieweg, S. 143-162.
- Huchler, Norbert/Pfeiffer, Sabine (Hg.) (2018): Industrie 4.0 konkret. Ungleichzeitige Entwicklungen, arbeitspolitische Einordnungen (= Schwerpunktheft der WSI-Mitteilungen 71/3), Baden-Baden: Nomos.
- Huchler, Norbert/Rhein, Philipp (2017): „Arbeitshandeln und der digitale Wandel von KMU. Die Rolle des Menschen und die Grenzen der Formalisierung 4.0“, in: Arbeit. Zeitschrift für Arbeitsforschung, Arbeitsgestaltung und Arbeitspolitik 26/3, S. 287-314.
- Huchler, Norbert/Voß, Günter G./Wehrich, Margit (2007): Soziale Mechanismen im Betrieb. Empirische und theoretische Analysen zur Entgrenzung und Subjektivierung von Arbeit, München/Mering: R. Hampp.
- IAB-Forschungsbericht (2016): Wirtschaft 4.0 und die Folgen für Arbeitsmarkt und Ökonomie. IAB-Forschungsbericht 13, <http://doku.iab.de/forschungsbericht/2016/fb1316.pdf> vom 11.11.2017.
- Ittermann, Peter/Niehaus, Jonathan/Hirsch-Kreinsen, Hartmut/Dregger, Johannes/ten Hompel, Michael (2016): Social Manufacturing and Logistics. Gestaltung von Arbeit in der digitalen Produktion und Logistik (= Soziologisches Arbeitspapier 47), Dortmund: Technische Universität Dortmund.
- Latour, Bruno (1987): Science in Action: How to Follow Scientists and Engineers Through Society, Milton Keynes: Open University Press.
- Lee, Edward A. (2008): Cyber Physical Systems: Design Challenges, University of California at Berkeley, Technical Report UCB/EECS-2008-8, 23.1.2018, <https://www2.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-8.pdf> vom 20.8.2018.
- Lindblom, Charles E. (1959): „The Science of Muddling-Through“, in: Public Administration Review 19, S. 79-88.

- Lobe, Adrian (2018): Die Gesellschaft der Metadaten, Süddeutsche online vom 31.7.2018, <https://www.sueddeutsche.de/digital/philosophie-die-gesellschaft-der-metadaten-1.4070474> vom 20.8.2018.
- McKinsey Global Institut (2017): Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation, <https://www.mckinsey.com/~/media/mckinsey/featured%20insights/Future%20of%20Organizations/What%20the%20future%20of%20work%20will%20mean%20for%20jobs%20skills%20and%20wages/MGI-Jobs-Lost-Jobs-Gained-Report-December-6-2017.ashx> vom 20.8.2018.
- Moldaschl, Manfred/Sauer, Dieter (2000): Internalisierung des Marktes – zur neuen Dialektik von Kooperation und Herrschaft. In: Heiner Minssen (Hg.), Begrenzte Entgrenzungen. Wandlungen von Organisation und Arbeit, Berlin: edition sigma, S. 205-224.
- Moldaschl, Manfred/Voß, Günter G. (Hg.) (2003): Subjektivierung von Arbeit, München/Mering: R. Hampp.
- Mühge, Gernot (2018): „Einzug der Rationalität in die Organisation? Digitale Systeme der Entscheidungsunterstützung in der Produktion“, in: Norbert Huchler/Sabine Pfeiffer (Hg.), Industrie 4.0 konkret. Ungleichzeitige Entwicklungen, arbeitspolitische Einordnungen (= Schwerpunktheft der WSI-Mitteilungen 71/3), S. 189-195.
- Neumer, Judith/Pfeiffer, Sabine (2015): „On the Importance of Experience-Based Work Action and Tacit Knowledge for Workplace and Social Innovation“, in: Warsaw Forum of Economic Sociology 6(11), S. 101-115.
- Neumer, Judith/Porschen-Hueck, Stephanie/Sauer, Stefan (2018): „Reflexive Scaling as a Way Towards Agile Organizations“, in: Journal of International Management Studies 18/2, S. 27-38.
- Neuweg, Georg H. (2015): Das Schweigen der Könner. Gesammelte Schriften zu implizitem Wissen, Münster/New York: Waxmann.
- Niehaus, Jonathan (2017): Mobile Assistenzsysteme für Industrie 4.0 – Gestaltungsoptionen zwischen Autonomie und Kontrolle, FGW-Publikation Digitalisierung von Arbeit 04, Düsseldorf: FGW, http://www.fgw-nrw.de/fileadmin/user_upload/FGW-Studie-I40-04-Niehaus-A1-web-komplett.pdf vom 28.8.2018.

- Perrow, Charles (1987): Normale Katastrophen – die unvermeidbaren Risiken der Großtechnik, Frankfurt a.M./New York: Campus.
- Pfeiffer, Sabine (2015): „Warum reden wir eigentlich über Industrie 4.0? Auf dem Weg zum digitalen Despotismus“, in: Mittelweg 36. Zeitschrift des Hamburger Instituts für Sozialforschung 6, S. 14-36.
- Pfeiffer, Sabine/Suphan, Anne (2015): Der Mensch kann Industrie 4.0. Kurzfassung zum AV-Index. Stuttgart: Universität Hohenheim. http://www.sabine-pfeiffer.de/files/downloads/2015_Mensch_kann_Industrie40.pdf vom 20.8.2018.
- Plattform Industrie 4.0 (Hg.) (2015): Whitepaper FuE-Themen. Industrie 4.0, 7.4.2015, <https://www.din.de/blob/67744/de1c706b159a6f1baceb95a6677ba497/whitepaper-fue-themen-data.pdf> vom 20.8.2018.
- Polanyi, Michael (1985): Implizites Wissen, Frankfurt a.M.: Suhrkamp.
- Porschen-Hueck, Stephanie/Huchler, Norbert (2016): „Offene Organisation: Anforderungen, Strategien, Kompetenzen“, in: Betriebskultur und Kompetenznutzung machen den Unterschied, Schwerpunkttheft PERSONALquarterly 86/02, S. 9-15.
- Promotorengruppe Kommunikation der Forschungsunion Wirtschaft – Wissenschaft (2013): Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0, Frankfurt a.M.: Promotorengruppe Kommunikation der Forschungsunion Wirtschaft – Wissenschaft, https://www.acatech.de/wp-content/uploads/2018/03/Abschlussbericht_Industrie4.0_barrierefrei.pdf vom 23.8.2018.
- Rammert, Werner (2009): „Hybride Handlungsträgerschaft: Ein soziotechnisches Modell verteilten Handelns“, in: Otthein Herzog/Thomas Schildhauer (Hg.), Intelligente Objekte, Berlin: Springer, S. 23-33.
- Rammert, Werner/Schulz-Schaeffer, Ingo (2002): Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik, Frankfurt a.M.: Campus.
- Robertson, Brian J. (2016): Holacracy: Ein revolutionäres Management-System für eine volatile Welt, München: Vahlen.

- Scheer, August-Wilhelm (2012): Industrie 4.0: Alter Wein in neuen Schläuchen? Scheer Group online, <http://www.august-wilhelm-scheer.com/2012/02/08/industrie-4-0-alter-wein-in-neuen-schlauchen/> vom 20.8.2018.
- Schröder, Christian (2017): Herausforderungen von Industrie 4.0 für den Mittelstand, Friedrich-Ebert-Stiftung, <http://library.fes.de/pdf-files/wiso/12277.pdf> vom 20.8.2018.
- Spath, Dieter/Ganschar, Oliver/Gerlach, Stefan/Hämmerle, Moritz/Krause, Tobias/Schlund, Sebastian (Hg.) (2013): Produktionsarbeit der Zukunft – Industrie 4.0, Stuttgart: Fraunhofer.
- Stich, Volker (2016): Digitalisierung der Wirtschaft – Wer ist hiervon nicht betroffen? Vortrag auf der MTM-Bundestagung „Vernetzt, mobil – und produktiv? Der Mensch in der Arbeitswelt 4.0“, 27.10.2016, Stuttgart.
- Suchman, Lucy A. (2007): Plans and Situated Actions. The Problem of Human-Machine Communication, Cambridge u.a.: Cambridge University Press.
- Sydow, Jörg (1985): Der soziotechnische Ansatz der Arbeits- und Organisationsgestaltung, Frankfurt a.M.: Campus.
- Weick, Karl E./Sutcliffe, Kathleen M. (2003): Das Unerwartbare managen. Wie Unternehmen aus Extremsituationen lernen, Stuttgart: Klett-Cotta.

Die Autonomie instrumentell genutzter Technik

Eine handlungstheoretische Analyse

Ingo Schulz-Schaeffer

1. EINLEITUNG

Die Auseinandersetzung mit Autonomen Systemen in der Arbeitswelt, die Gegenstand der Tagung war, aus welcher der vorliegende Sammelband hervorgegangen ist, führt zu einer Vielzahl konkreter Fragen, die etwa neue Qualifikationsanforderungen an die Beschäftigten zum Gegenstand haben, sich wandelnde Arbeitsbedingungen oder Veränderungen in der Arbeitsorganisation. Sie gibt darüber hinaus auch Anlass, grundsätzlicher über das Zusammenwirken menschlich und technisch durchgeführter Arbeit nachzudenken. Zu dieser grundsätzlicheren Frage soll der vorliegende Artikel einen Beitrag leisten.

Der Beitrag besteht in einer handlungstheoretischen Analyse der Handlungsbeiträge technischer Artefakte zu Handlungszusammenhängen, die im Zusammenwirken menschlichen Tuns und technischen Wirkens realisiert werden. Anders als es das Thema der Tagung und des Sammelbandes erwarten lässt, beginnt meine Analyse jedoch nicht mit der Betrachtung des Zusammenwirkens von Mensch und Technik im Fall besonders autonomer Technik, sondern am anderen Ende des Spektrums. Meine argumentative Strategie ist, bei der Frage nach der

Eigenständigkeit gegenständlicher Technik an dem Fall der im denkbar geringsten Umfang autonomen Technik anzusetzen und von dort aus schrittweise zu Fällen größerer technischer Autonomie vorzustoßen. Dabei werde ich eine strikt relationale Perspektive einnehmen. Autonomie ist aus dieser Perspektive keine Eigenschaft, die sich direkt aus der Funktionsweise der technischen Artefakte ableiten lässt, sondern ein Merkmal der Verteilung von Handlungskompetenzen innerhalb von Handlungszusammenhängen, die im Zusammenwirken menschlichen Handelns und technischen Wirkens durchgeführt werden. Damit soll selbstverständlich nicht geleugnet werden, dass es für die Frage, wie selbstbestimmt und selbsttätig technische Artefakte in solchen Handlungszusammenhängen agieren können, auf deren technische Funktionalitäten ankommt. Eine entsprechende Analyse habe ich selbst an anderer Stelle vorgenommen (vgl. Schulz-Schaeffer 2008a). Hier geht es mir allerdings um die Frage der Autonomie von Technik im Rahmen ihres Einsatzes als Arbeitsmittel, oder allgemeiner ausgedrückt: im Rahmen ihrer instrumentellen Nutzung im zweckgerichteten Handeln. Und diese Frage erfordert die relationale Perspektive.

Im Folgenden werde ich zuerst drei sehr grundlegende Formen der instrumentellen Nutzung von Technik im zweckgerichteten Handeln unterscheiden (Kapitel 2), mit denen sich verschiedene Formen der Delegation von Handlungskompetenz verbinden. Die betreffende Handlungskompetenz fasse ich als spezifische Handlungssubjektqualitäten und unterscheide zu diesem Zweck zwischen einer effektiven, einer regulativen und einer intentionalen Handlungsdimension (Kapitel 3). Diese Konzeption bildet das Grundgerüst meiner handlungstheoretischen Analyse. Diese Analyse werde ich in den Kapiteln 4 bis 7 für die drei Grundformen der instrumentellen Nutzung von Technik durchführen. Dazu betrachte ich jeweils für die drei Handlungsdimensionen, in welcher Weise sich die Gesamthandlung und die Verteilung der Handlungsbeiträge ändert, wenn Teilstücke aus ihr an technische Artefakte delegiert werden. Der Beitrag schließt mit einer kurzen Zusammenfassung.

2. DREI FORMEN DER INSTRUMENTELLEN NUTZUNG VON TECHNIK IM ZWECKGERICHTETEN HANDELN

Die Konstellation des zweckrationalen Handelns, das sich technischer Wirkungen instrumentell bedient, ist diejenige Mensch-Technik-Konstellation, in der sich die denkbar asymmetrischsten Formen der Verbindung von menschlichem Handeln und technischem Wirken finden. Hier sind diejenigen Wirkungszusammenhänge aus menschlichem Handeln und technischem Wirken angesiedelt, die durch die denkbar geringste Autonomie technischer Artefakte und die denkbar umfassendste Steuerung und Kontrolle durch menschliches Handeln gekennzeichnet sind.

Dies ist zum einen durch diesen Typus des Handelns begründet, zum anderen durch die Art und Weise der Bezugnahme auf Technik: Das zweckrationale Handeln ist von allen idealtypischen Formen des Handelns diejenige, bei der der Entwurf und die Durchführung der Handlung am vollständigsten durch eine konkret gefasste Intention des oder der Handelnden gesteuert wird: durch ein Handlungsziel, von dem sich die Schritte der Handlungsdurchführung ableiten. Die instrumentelle Nutzung von Technik ist von allen Formen der Techniknutzung diejenige, die am ausdrücklichsten und umfassendsten darauf ausgerichtet ist, den Beitrag des technischen Wirkens zu dem jeweiligen Geschehen von menschlichen Handlungszielen abhängig zu machen.

Es gibt allerdings unterschiedliche Formen der instrumentellen Bezugnahme auf Technik. Dies ist für unsere Fragestellung deshalb bedeutsam, weil mit ihnen unterschiedliche Grade menschlicher Kontrolle und technischer Autonomie einhergehen. Es ist zweckmäßig, zwischen den folgenden drei Formen der instrumentellen Nutzung von Technik zu unterscheiden:

(1) Arbeitersparnis: Handlungsvollzüge werden an technische Artefakte delegiert, um menschliche Arbeit zu sparen. Dieses instrumentelle Motiv der Techniknutzung tritt bei solchen Tätigkeiten am deutlichsten zutage, bei denen es für die Qualität des Handlungsergebnisses

keine wesentliche Rolle spielt, ob sie an Technik delegiert wird oder nicht (z.B. Geschirr spülen).

(2) Bessere Handlungsdurchführung: Vielfach geht es bei der Techniknutzung nicht nur darum, menschliche Arbeitskraft zu sparen, sondern darum, das Handlungsziel auch in anderen Hinsichten besser realisieren zu können, als es mit den eigenen Händen möglich wäre (z.B. schneller, präziser, fehlerfreier usw.).

(3) Handlungsermöglichung: Oder aber es geht darüber hinaus darum, völlig neue Handlungsmöglichkeiten zu gewinnen. Diese dritte Form tritt im Laufe der Entwicklung der wissenschaftlich-technischen Zivilisation zunehmend in den Vordergrund: Mehr und mehr sind technische Neuerungen Bestandteile von Handlungen, die durch diese Techniken überhaupt erst ermöglicht werden.

Diese Dreiteilung ist eine analytische Unterscheidung. Im empirischen Fall überlagern sich immer wieder mehrere dieser Formen. Zudem sind die Grenzen zwischen ihnen häufig fließend und können sich im Zeitverlauf verschieben. Ihr Wert als analytische Unterscheidung besteht darin, dass sie es erlaubt, unterschiedliche Grundmuster der Einbeziehung von Handlungsmitteln in die Handlungsplanung und -durchführung zu unterscheiden, die für die Frage der Verteilung von Handlungssubjektqualitäten folgenreich ist.

3. DREI HANDLUNGSDIMENSIONEN

Handlungstheoretisch betrachtet sind Handlungen Verhaltenssequenzen, die der Realisierung bestimmter Ziele dienen (vgl. Parsons 1968 [1937]: 44) und die daran in ihrer Durchführung orientiert sind. Handeln ist dementsprechend ein Verhalten, dessen Entwurf den (bewusst oder stillschweigend in Anschlag gebrachten) Intentionen von Akteur/-innen entspringt und dessen Durchführung ihrer intentionalen Steuerung und Kontrolle unterliegt (vgl. z.B. Luckmann 1992: 48-52).

Ausgehend von diesem in der Soziologie und den Sozialwissenschaften weit verbreiteten Handlungsbegriff lassen sich Akteur/-innen

durch drei Merkmale charakterisieren: Sie sind (1) als effektive Handlungssubjekte fähig, verändernde Wirksamkeit auszuüben, sind also in der Lage, die Veränderungen in Raum und Zeit herbeizuführen, die erforderlich sind, um das betreffende Handlungsziel zu erreichen (*effektive Handlungsdimension*). Sie sind (2) regulative Handlungssubjekte, insofern sie die Steuerungsgewalt und Kontrolle über die Handlungsdurchführung besitzen (*regulative Handlungsdimension*); und sie sind (3) intentionale Handlungssubjekte, das heißt, es sind ihre Zielsetzungen, von denen sich die Schritte der Handlungsdurchführung ableiten (*intentionale Handlungsdimension*).

Zweckrational handelt Max Weber (1972 [1922]: 13) zufolge, „wer sein Handeln nach Zweck, Mitteln und Nebenfolgen orientiert und dabei sowohl die Mittel gegen die Zwecke, wie die Zwecke gegen die Nebenfolgen, wie endlich auch die verschiedenen möglichen Zwecke gegeneinander rational *abwägt*“ (Herv. i.O.). Das zweckrationale Handeln repräsentiert mithin den idealtypischen Fall eines rational konsistenten Zusammenhanges zwischen den drei Handlungsdimensionen – ein idealtypischer Fall, weil, wie bereits Weber notiert, reine Zweckrationalität „ein im wesentlichen konstruktiver Grenzfall“ (ebd.) ist.

4. DIE DELEGATION VON TEILSTÜCKEN BESTEHENDER HANDLUNGEN AN TECHNIK

Zuerst betrachte ich den Fall, in dem technische Artefakte eingesetzt werden, um Teilstücke bereits bestehender Zwecktätigkeiten zu übernehmen. Ein Beispiel ist die Kaffeemaschine für Filterkaffee. Die vorgängig existierende Handlung der Zubereitung von Filterkaffee besteht darin, Wasser zu kochen, eine Filtertüte in einen Kaffeefilter zu stecken, die gewünschte Menge gemahlene Kaffees einzufüllen, den Filter auf eine Kaffeekanne zu stellen und dann nach und nach – um zu vermeiden, dass der Filter überläuft – heißes Wasser aufzugießen, solange bis die gewünschte Menge gefilterten Kaffees in die Kanne durchgelaufen ist. Die Kaffeemaschine übernimmt aus dieser Hand-

lung die Teilhandlungen des Wasserkochens und des portionierten Aufgießens von Wasser. Technische Artefakte, die in dieser Weise eingesetzt werden, übernehmen Tätigkeiten, die menschliche Akteur/-innen selbst genauso gut verrichten könnten. Ihr primärer instrumenteller Nutzen besteht mithin in der Arbeitersparnis, darin also, dass sie Menschen Arbeit abnehmen, die sie andernfalls selbst verrichten müssten.

Für den vorliegenden Zusammenhang ist diese Form der Handlungsdelegation an Technik deshalb interessant, weil sie diejenige Form der instrumentellen Bezugnahme auf Technik im zweckrationalen Handeln darstellt, die – wie die handlungstheoretische Analyse zeigt – durch die geringste Autonomie technischer Artefakte und die umfassendste Steuerung und Kontrolle durch menschliches Handeln gekennzeichnet sind.

Zweckrationales Handeln – so wie alles zielgerichtete Handeln – ist ein Verhalten, das seinen Sinn und seine Bedeutung vom Handlungsziel her gewinnt. Akteur/-innen, die bestimmte Handlungen zum ersten Mal planen, entwerfen ihre Handlungen, indem sie sich ausgehend vom Handlungsziel überlegen, welche Handlungsschritte in welcher Reihenfolge erforderlich sind, um zu diesem Ziel zu gelangen. Die tatsächliche Handlung erfolgt dann als ein an diesem Entwurf orientiertes Handeln. Diese grundlegende Sinnstruktur haben auch wiederholt durchgeführte Handlungen wie das Kaffeekochen. Sofern sie durchgeführt werden, um das betreffende Handlungsziel zu realisieren, ist auch bei ihnen der Sinn der einzelnen Handlungsschritte von diesem Handlungsziel abgeleitet. Der Unterschied ist nur, dass der aus den einzelnen Handlungsschritten bestehende Handlungsentwurf bereits als vorgefertigtes Wissen im persönlichen Erfahrungswissen und/oder Routinewissen vorliegt und dann häufig auch bereits in der einen oder anderen Weise Bestandteil des gesellschaftlichen Wissensvorrates ist (vgl. Luckmann 1992: 48-67).

In dem hier betrachteten Fall werden nun einzelne Teilhandlungen eines solchen, bereits vorgefertigten und im gesellschaftlichen Wissensvorrat vorhandenen Handlungsentwurfs technisch objektiviert, also

in Gestalt technisch realisierter Wirkungszusammenhänge bereitgestellt, die sich bei Bedarf abrufen lassen – also dann, wenn die Handlungsdurchführung an dem Punkt angekommen ist, an dem der Handlungsentwurf die betreffenden Teilhandlungen vorsieht. Im Beispiel der Kaffeemaschine ist das sehr deutlich: Die Kaffeemaschine wird genau an dem Punkt der Handlungsdurchführung eingeschaltet, an dem die an sie delegierte Tätigkeit des Wasseraufgießens zuvor von menschlichen Akteur/-innen durchgeführt wurde.

Aus diesem direkten Substitutionsverhältnis folgt, dass sich an der Sinnstruktur der Gesamthandlung und ihrer Teilstücke nichts Grundsätzliches ändert. In den Worten von Hans Freyer: Daraus, dass „ein bloßes Teilstück einer Gesamthandlung den Sinngehalt der Objektivation bildet“, ergibt sich, dass auch das technisch bereitgestellte Teilstück

„seinem Sinn nach wesentlich Teil [bleibt], es genügt sich nicht selbst, sondern will dem Ganzen der Handlung dienstbar wieder eingefügt werden. Es weist immer außer sich, es ist nie Gebilde mit in sich geschlossenem Bedeutungsgesamt. Sondern es fordert eine Erfüllung seiner Sinnintentionen durch hinzutretende Akte der Verwendung.“ (Freyer 1966 [1934]: 61)

In welcher Weise ändert sich die Handlung also, wenn Teilstücke aus ihr in der dargestellten Weise an technische Artefakte delegiert werden? Aus den eben angestellten Überlegungen folgt, dass sich in der intentionalen Handlungsdimension nichts ändern sollte: Der Handlungsentwurf und die durch ihn bestimmten Teilstücke der Handlung beziehen ihren Sinn aus dem Handlungsziel. Am Handlungsziel hat sich nichts geändert und auch der Handlungsentwurf ist in seiner Struktur im Wesentlichen gleich geblieben. Nur werden jetzt eben einzelne Teilstücke der Handlung von technischen Artefakten übernommen. Das heißt, die gleiche zielsetzende Intention, die zuvor dem menschlichen Tun seinen Sinn verliehen hat, bestimmt jetzt die Bedeutung der technisch realisierten Teilstücke der Handlung.

In der effektiven Handlungsdimension verändert sich die Situation dagegen. Die verändernde Wirksamkeit wird jetzt zum Teil von der Technik ausgeübt und nicht von menschlichen Akteur/-innen. Damit geht zwangsläufig auch eine Veränderung in der regulativen Handlungsdimension einher. Der Akteur, der den Kaffee mit seinen eigenen Händen aufgießt, hat noch die vollständige Steuerungsgewalt beispielsweise darüber zu verhindern, dass der Kaffeefilter überläuft. Wird diese Tätigkeit an die Kaffeemaschine delegiert, so geht zugleich auch die Steuerung dieses Geschehens an sie über. Wir können folglich sagen, dass ein Stück des Handlungssinns nun in technisch objektivierter Form vorliegt. Der vom Handlungsziel über den Handlungsentwurf an die Teilstücke der Handlung übertragene Handlungssinn liegt dem technischen Ablauf genauso zugrunde wie dem menschlichen Handeln, an dessen Stelle es tritt. Er liegt jetzt als technisch objektivierter Handlungssinn vor.¹

Die handlungstheoretische Analyse zeigt, dass die hier betrachtete Form der Einbeziehung technischer Wirksamkeit in menschliches Handeln in der intentionalen Handlungsdimension auf weitgehend unverändert bleibenden Handlungszielen beruht und in der regulativen Handlungsdimension auf Handlungsentwürfen, die in ihrer Struktur weitgehend unverändert bleiben. Dies begründet das Urteil, dass es sich hier um die asymmetrischste Form instrumenteller Technikenutzung handelt, um diejenige, die durch die denkbar geringste Autonomie technischer Artefakte und die denkbar umfassendste Steuerung und Kontrolle durch menschliches Handeln gekennzeichnet ist. Zugleich zeigt die Analyse aber auch, dass die technischen Artefakte bereits hier einen Teil des sinnhaften Handelns und dessen sinnhafter Steuerung übernehmen.

1 Zur Frage der technischen Objektivierung von Handlungssinn vgl. auch Schulz-Schaeffer 2007: 256-258, 441-442, 483-485. Zum zugrunde liegenden Verständnis von Sinn als erzeugter Selektivität vgl. Luhmann 1984: 101; Schulz-Schaeffer 2007: 33-39.

5. DER EINFLUSS DER TECHNISCHEN OBJEKTIVIERUNG VON HANDLUNGSSINN AUF DIE ZIELSETZENDEN INTENTIONEN

Das zuvor gezeichnete Bild wird etwas komplexer, wenn man einbezieht, dass die Tatsache der technischen Objektivierung von Handlungsteilstücken Rückwirkungen auf die Ausrichtung und Wahl von Handlungszielen haben kann und vielfach auch hat. Indirekt ergibt sich dadurch dann doch bereits bei dieser Form der instrumentellen Nutzung von Technik ein Einfluss auf der Ebene der intentionalen Handlungsdimension. Dies wird deutlich, wenn man die auf ein konkretes Handlungsziel hin entworfene Handlung mit einer technisch unterstützten Handlung entsprechender Art vergleicht. Ich verwende dazu das Beispiel, mit dem Freyer seine zuvor zitierten Überlegungen illustriert:

„Sagen wir: ich will Wasser trinken. Ich muß erst die Quelle suchen, dann mich bücken, dann die Hand hohl machen, Wasser auffangen usw. Aus diesem Handlungszusammenhang wird (wie das genetisch zugeht, fragen wir nicht) ein einzelnes Teilstück ‚Herstellung einer Höhlung mit dichtem Boden‘ zur Form objektiviert, es wird [...] aus der bloßen Aktion abgelöst und (in unserm Falle auch materiell) verfestigt.“ (Freyer 1966 [1934]: 61)

Über die technisch objektivierten Teilstücke von Zwecktätigkeiten sagt Freyer:

„Liegen dann die Formen fertig vor, so findet der Handlungszusammenhang ein Stück seiner selbst vorgetan. Er läuft durch das Gerät hindurch, er läuft das betreffende Stück gleichsam auf Schienen, während er sich sonst seinen Weg frei durch die gegenständliche Welt hindurch suchen muß.“ (Ebd.: 61-62)

Während also der durstige Wanderer, der ohne technische Unterstützung aus der Quelle trinkt, seine Hände in der Situation gerade so formen kann, wie es den konkreten Gegebenheiten am besten entspricht – aus ihnen beispielsweise eine schmalere Schale bildet, wenn das Was-

ser in einer engen Felsspalte fließt –, ist er bei Benutzung der Schale zwar von jener Arbeit entlastet, dafür aber auf die Form festgelegt, die durch die materiell objektivierte Gestalt des Artefaktes vorgegeben ist. Der in dem technischen Artefakt objektivierte Handlungsschritt ist mithin nicht in der gleichen Unmittelbarkeit aus dem konkreten Handlungsziel des betreffenden menschlichen Akteurs abgeleitet, wie dies bei seinen Körperbewegungen möglich ist.

Diese geringere Unmittelbarkeit hat ihre Ursache darin, dass Schalen nicht dazu da sind, um in einer konkreten Situation ein Teilstück einer konkreten Handlung zu bilden, sondern um diesen Beitrag zu vielen Handlungen ähnlicher Art beisteuern zu können. Auch insgesamt betrachtet ist es die Ausnahme, dass ein technisches Artefakt aus einem konkreten Handlungsziel abgeleitet ist und für die Nutzung in einer konkreten Handlungssituation konstruiert wird. Im Normalfall wird instrumentelle Technik vielmehr entwickelt und hergestellt, um in bestimmten Situationen immer wieder für bestimmte Zwecke genutzt werden zu können. Das Merkmal der wiederholten Nutzbarkeit ist geradezu eines der Definitionsmerkmale von Technik (vgl. Schulz-Schaeffer 2008b: 445-447). Dieses Merkmal ergibt sich zumeist bereits aus dem Umstand, dass die Entwicklung und Herstellung der betreffenden technischen Arrangements mit einem Aufwand verbunden ist, der sich unter Maßgabe zweckgerichteten Handelns erst bei wiederholter Nutzung lohnt.

Für den vorliegenden Diskussionszusammenhang ist die entscheidende Konsequenz aus diesem Sachverhalt, dass die Sinnzusammenhänge, von denen instrumentelle Techniken ihre Bedeutung zugewiesen bekommen, nicht die konkreten Handlungen sind, mit denen konkrete Personen ihre individuellen Handlungsziele zu realisieren trachten. Es sind vielmehr typische Handlungen, mit denen typische Akteur/-innen die in typischen Situationen normalerweise zu erwartenden Handlungsziele verfolgen. Die im technischen Artefakt objektivierte Sinnintention ist also die Bereitstellung typischer Schritte von Handlungsvollzügen für typische Handlungen (vgl. Schulz-Schaeffer 2008c: 711-712). Wenn also Technik eingesetzt wird, um Teilstücke bereits

bestehender Handlungen zu übernehmen, dann sind es häufig Handlungen, deren Entwürfe bereits als vorgefertigte Handlungsmuster im gesellschaftlichen Wissensvorrat bereitliegen. Und dann gilt: Was jetzt technisch objektiviert wird, lag zuvor bereits in kulturell objektivierter Form vor.²

Der Akteur, der eine instrumentelle Technik nutzen möchte, die ihrerseits ein typisches Teilstück einer zugehörigen typischen Handlung durchführt, wird dadurch – zumindest in einem gewissen Umfang – gezwungen, die typische Handlung durchzuführen, zu der das technisch objektivierte Teilstück passt. In dem Maße, in dem dies der Fall ist, verfolgt er dann aber auch das Handlungsziel dieser typischen Handlung. Der Akteur muss mithin seine gegebenenfalls davon abweichende Handlungsabsicht der entsprechenden typischen Handlungsabsicht anpassen, um seine Handlung als eine Instanz der betreffenden typischen Handlung durchführen zu können.

Muss man aus diesen Überlegungen den Schluss ziehen, dass Technik bereits bei dieser asymmetrischsten Form ihrer instrumentellen Nutzung in der intentionalen Handlungsdimension wirksam ist? Die Antwort hängt von der Differenz zwischen dem konkreten und dem typischen Handlungsziel ab. In dem Maße, in dem wir gewohnt sind, unsere Handlungen auf der Grundlage der vorliegenden gesellschaftlichen Sinnmuster zu bilden und uns das typische Handlungsziel und das zugehörige Handlungsmuster gleichermaßen zu eigen machen, entsteht keine Differenz oder bleibt sie unmerklich. Aus diesem Grund werden vermutlich wenige Nutzer/-innen von Kaffeemaschinen den Eindruck haben, dass deren Benutzung sie nötigt, ihre eigentlichen Handlungsziele modifizieren zu müssen. Erst die Abweichung von den

2 Das Vorliegen kulturell objektivierter Handlungsmuster ist zugleich eine zentrale Ermöglichungsbedingung für die Delegation von Teilhandlungen an Technik. Wo die Handlung bereits zuvor durch ein kulturell verfestigtes Skript angeleitet wurde, ist es einfacher, ein Teilstück herauszulösen und an einen technischen Ablauf zu delegieren (vgl. Rammert/Schulz-Schaeffer 2002: 44–45).

Zielen und Vorgehensweisen, die qua technischer Objektivierung in die technischen Artefakte eingeschrieben sind, lässt die Differenz zutreten. Und ob sie dann mit Anpassung an das Gerät beantwortet wird oder mit Anpassung des Geräts, ist eine empirische Frage (vgl. Akrich 1992).

6. HANDLUNGSDELEGATION AN ARTEFAKTE MIT SPEZIALKOMPETENZ

Vielfach geht es beim instrumentellen Einsatz von Technik im zweckgerichteten Handeln nicht nur darum, menschliche Arbeitskraft zu sparen, sondern darum, das Handlungsziel auch in anderen Hinsichten besser realisieren zu können, als es mit den eigenen Händen möglich wäre (z.B. schneller, präziser, fehlerfreier usw.). Auch hier gibt es eine vorgängige Form menschlichen Handelns, das den Vergleichsfall bildet: die Delegation der Durchführung von Handlungen oder von Teilen daraus an menschliche Expert/-innen im Rahmen arbeitsteiliger Spezialisierung. So wie also die mit den eigenen Händen durchgeführte Handlung die Hintergrundfolie bildet, um besser zu sehen, in welchem Umfang und in welchen Handlungsdimensionen technische Artefakte zu Handlungssubjekten werden, wenn Teilstücke aus bereits etablierten Handlungsabläufen an technische Artefakte delegiert werden, so bildet Handlungsdelegation im Rahmen arbeitsteiliger Spezialisierung die Hintergrundfolie für die entsprechende Analyse der Handlungsbeteiligung von technischen Artefakten, die Tätigkeiten in irgendeiner Weise besser durchführen können als ihre Nutzer/-innen.

Wie im zuvor betrachteten Fall wird ein in dieser Weise genutztes technisches Artefakt in der effektiven Handlungsdimension zum Handlungssubjekt der an es delegierten Handlungsvollzüge. In der regulativen Handlungsdimension zeigt sich dagegen ein wesentlicher Unterschied zum zuvor betrachteten Fall. Um sich dies zu verdeutlichen, ist es hilfreich zwischen ‚etwas steuern‘ und ‚etwas unter Kontrolle haben‘ zu unterscheiden. Etwas zu steuern ist der Vorgang, mit dem ein

Vollzug an einem Plan orientiert wird, in unserem Fall also: die Handlungsdurchführung am Handlungsentwurf. Etwas unter Kontrolle zu haben bedeutet dagegen, die Übereinstimmung oder Abweichung von Plan und Realisierung sehen zu können und im Fall einer Abweichung eingreifen zu können.

Sowohl im Fall der arbeitssparenden Delegation von Tätigkeiten an Technik, die deren Nutzer/-innen auch selbst durchführen könnten, wie auch im Fall der Delegation gekonnter Tätigkeiten geht mit der effektiven Realisierung der delegierten Teilhandlungen zugleich auch deren Steuerung an die technischen Artefakte über. Unterschiede bestehen dagegen bezüglich der Kontrolle des delegierten Geschehens. Die Kontrolle verlagert sich, zumindest im Prinzip, am wenigsten bei Tätigkeiten, die menschliche Akteur/-innen genauso gut selbst erledigen könnten und die nur aus Gründen der Arbeitersparnis technisch delegiert werden. Wenn man beispielsweise das Kaffeeaufgießen an die Kaffeemaschine delegiert, gibt man zwar die Steuerung des aktuellen Geschehens an die Maschine ab, man kann es aber weiter unter der eigenen Kontrolle halten. Man weiß, wie diese Tätigkeit geht, und kann es dementsprechend sofort bemerken, wenn etwas schief läuft. Die Situation ändert sich deutlich, wenn eine Technik deshalb eingesetzt wird, weil sie die betreffende Tätigkeit in bestimmten Hinsichten besser durchzuführen verspricht, als es der menschliche Akteur könnte.

Die Schwierigkeit, ein Artefakt zu kontrollieren, das etwas besser kann als man selbst, besteht darin, dass man dazu die Kompetenz erwerben müsste, die für diesen Unterschied verantwortlich ist. Diese Kompetenz nicht besitzen zu müssen, sondern das betreffende Teilstück der Handlung an das diesbezüglich kompetentere Artefakt delegieren zu können, macht aber gerade dessen instrumentellen Nutzen aus – ganz abgesehen davon, dass es vielfach um Kompetenzen geht, die die menschlichen Fähigkeiten grundsätzlich übersteigen. Dazu bedarf es nicht einmal sonderlich komplexer Technologien. Eine einfache handelsübliche Banknotenzählmaschine beispielsweise kann tausend Banknoten pro Minute zählen. Das kann kein Mensch. Sicherlich kann man testen, ob die Maschine erwartungsgemäß funktioniert. Das ge-

schieht natürlich auch. Daraus lässt sich die Erwartung ableiten, dass sie mit einer bestimmten Wahrscheinlichkeit auch zukünftig die gewünschten Handlungsbeiträge liefern wird. Im konkreten Handlungsvollzug selbst können die menschlichen Akteur/-innen dies jedoch nicht kontrollieren. Sie müssen darauf vertrauen, dass die Maschine den Ablauf in ihrem Sinne steuert und kontrolliert.³

Wenn wir eine Banknotenzählmaschine benutzen, delegieren wir die Durchführungssteuerung an Verhaltenseigenschaften, die dem technischen Artefakt qua seiner objektivierten Gestalt fest eingeschrieben sind. Dies ist auch bei vielen anderen technischen Arrangements so, die technisch nicht sehr kompliziert sein müssen, deren Wirken menschlichem Handeln aber dennoch überlegen ist. So etwa auch bei dem automatischen Türschließer, dessen unnachgiebiges Pflichtbewusstsein Bruno Latour (1992: 232) nicht genug loben kann. Bei Techniken dieser Art ist das Vertrauen in sie darin gegründet, dass die Artefakte im ordnungsgemäßen Zustand gar nicht anders können, als sich gemäß der in ihnen eingeschriebenen Verhaltenseigenschaften zu verhalten. Daraus speist sich auch die Gewissheit, dass die delegierte Durchführungssteuerung im Dienst der vorgegebenen Handlungsziele steht.

Wie aber sieht es aus, wenn die technisch delegierte Durchführungssteuerung den Geräten nicht fest eingeschrieben ist, sondern auf informationsabhängigen Entscheidungsprozessen beruht (vgl. Schulz-Schaeffer 2008a: 34-40, 44). Navigationsgeräte und Routenplaner sind Beispiele hierfür. Zumeist gibt es verschiedene Wege von A nach B. Ein Navigationsgerät, dem man die Aufgabe überträgt, einen auf dem besten Weg von einem Startpunkt zu einem Zielpunkt zu leiten, muss also die Vor- und Nachteile der alternativen Wegstrecken gegeneinander abwägen. Heutige Navigationssysteme verwenden dafür Kartenmaterial, das nicht nur Entfernungsdaten und allgemeine Informationen über Verkehrsführungen und Geschwindigkeitsbegrenzungen enthält,

3 Vgl. weiterführend zum Thema des Vertrauens in Technik Schulz-Schaeffer 2000: 349-363; Wagner 1994.

sondern darüber hinaus auch kontinuierlich aktualisierte Informationen etwa über geänderte Geschwindigkeitsbegrenzungen, gesperrte Straßen oder geänderte Abbiegeverbote und zudem in Echtzeit eingespielte Verkehrsinformationen über Verkehrsbehinderungen. Es kommt vor, dass ein solches Gerät einem einen anderen Weg weist, als man selbst gewählt hätte. Wenn das Gerät eine gute Software besitzt und die Informationen, die es einbezieht, hinreichend vollständig und aktuell sind, gibt es dafür nachvollziehbare Gründe. Der Grund ist dann beispielsweise, dass das Gerät ein besseres Wissen und Einschätzungsvermögen darüber hat, wieviel Fahrzeit die betreffenden Strecken tatsächlich erfordern, oder dass es Informationen über Verkehrsbehinderungen einbezogen hat, die einem selbst fehlten.

Techniken dieser Art übernehmen und steuern nicht lediglich ein in der Art und Weise seiner Durchführung bereits festgelegtes Teilstück von Zwecktätigkeiten. Sie übernehmen vielmehr auch dessen Planung.⁴ Die Planung eines Teils des Handlungsentwurfs der betreffenden Handlung erfolgt damit auf der Grundlage von Sinngehalten, über die das technische Artefakt verfügt, aber nicht sein menschlicher Nutzer. Das Artefakt organisiert und steuert den ihm übertragenen Teil der Handlung in einer anderen Weise als sein Nutzer, das eigentliche intentionale Subjekt der Handlung, es täte. Es trägt dadurch dessen Handlungsintentionen aber besser Rechnung, als jener selbst es könnte.

Dies ist ein eigentümliches und erklärungsbedürftiges Auseinandertreten von regulativem und intentionalem Handlungssinn, jedenfalls dann, wenn man handlungstheoretisch von der Grundannahme ausgeht, dass es die Intentionen des Handelnden sind, von denen her der Handlungsentwurf seinen Sinn bezieht. Die hier betrachtete Form techni-

4 Bei heutigen Navigationsgeräten für den Straßenverkehr beschränkt sich die Tätigkeit der Technik auf die Planung, während die Durchführung bei dem Autofahrer/der Autofahrerin bleibt, der/die den Anweisungen des Geräts folgen kann oder auch nicht. Technische Systeme wie die automatische Einparkhilfe dagegen übernehmen Planung und Durchführung des betreffenden Teilstücks der Handlung.

scher Delegation steht in deutlichem Gegensatz zu der dieser Grundannahme verpflichteten Auffassung von Alfred Schütz und Thomas Luckmann (1984: 15), dass „der Handelnde [...] die letzte Instanz [ist], die angehört werden muß, wenn es festzustellen gilt, ob in einem vorliegenden Fall gehandelt wird oder nicht. Nur er weiß, woraufhin – falls überhaupt auf etwas – das Geschehen entworfen wurde.“

Für die Erklärung dieses Auseinandertretens von intentionalem und regulativem Handlungssinn ist es hilfreich, den Bereich der technischen Delegation kurz zu verlassen und sich dem Handeln im Auftrag und im Interesse anderer zuzuwenden, wie es in arbeitsteiligen Beziehungen zwischen menschlichen Akteur/-innen gang und gäbe ist. Mit der arbeitsteiligen Organisation moderner Gesellschaften geht eine weitgehende Ausdifferenzierung spezialisierter körperlicher Fertigkeiten und kognitiver Fähigkeiten einher, die sich häufig nur durch jahrelange Ausbildung und Praxis erwerben lassen. Für sehr viele Tätigkeitsbereiche gibt es dementsprechend Fachleute und Expert/-innen, die die jeweiligen Tätigkeiten besser können als der Durchschnittsbürger/die Durchschnittsbürgerin und deren Fähigkeiten sich nicht umstandslos aneignen lassen. Bei der Delegation von Handlungsvollzügen an Fachleute und Expert/-innen ist es aus diesem Grund vielfach nicht nur ganz normal, mit ihrer Durchführung zugleich auch ihre Planung zu delegieren, sondern dies ist im Interesse der eigenen Handlungsziele geradezu geboten. Johannes Weiß, der die mit besonders wichtigen Handlungsinteressen verbundenen Formen delegierten Handelns unter dem Begriff der Stellvertretung in den Blick nimmt, zitiert als Buchmotto einen amerikanischen Richterspruch, der dies markant auf den Punkt bringt: „Wer sich vor Gericht selbst vertritt, hat einen Narren zum Anwalt.“ (Weiß 1998: 13)

Worin aber besteht die Kompetenz anderer, die es ihnen ermöglicht, meine eigenen Handlungsabsichten besser in die Tat umzusetzen, als ich selbst es könnte, sodass man im Zweifelsfall sie und nicht mich fragen müsste, was der Sinn des betreffenden Teilstücks der Handlung ist? Die Antwort auf diese Frage weist auf die obigen Überlegen über typische Handlungen und deren typische Teilstücke zurück. Ein kom-

petenter Rechtsanwalt weiß auf der Grundlage seiner Kenntnis der Gesetzeslage und der Rechtsprechung zu Fällen, die dem meinen ähneln, welche Verteidigungsstrategie in meinem Fall den besten Erfolg verspricht. Eine kompetente Ärztin weiß auf der Grundlage ihres medizinischen Wissens und ihrer bisherigen Erfahrung mit der Krankheit, die mich plagt, welche Behandlung in meinem Fall angebracht ist. Keiner dieser Expert/-innen muss meinen konkreten Fall kennen, um die für meinen Fall erforderliche Handlungskompetenz zu erlangen. Ihre Kompetenz besteht vielmehr in der Spezialisierung auf die Bewältigung typischer Handlungssituationen bestimmter Art. Diese Kompetenz kommt im konkreten Fall zum Tragen, wenn dieser sich als ein Anwendungsfall einer jener typischen Situationen fassen lässt.

Die Ausdifferenzierung spezialisierter Handlungskompetenz erklärt das Auseinandertreten von intentionalem und regulativem Handlungssinn. Wer Expert/-innen beauftragt, im eigenen Interesse und Auftrag zu handeln, akzeptiert damit bewusst oder stillschweigend, dass das delegierte Handeln zu einer Fallgruppe von Handlungen gehört, über die diese Expert/-innen besser Bescheid wissen als man selbst. Im zweckgerichteten Interesse an der Realisierung der eigenen Handlungsziele ist es unter diesen Umständen die beste Strategie, ihnen die Kontrolle über die Planung und Durchführung des Handelns zu überlassen.

Die Delegation von Handlungsentscheidungen an ‚intelligente‘ Technik folgt derselben Rationalität. Auch hier ist die überlegene Problemlösungsfähigkeit bezogen auf Handlungen bestimmter Art der Grund, sich lieber auf die Geräte zu verlassen, als auf die eigene Handlungskompetenz. In beiden Fällen beruht die Handlungsdelegation auf einem Vertrauen in die Richtigkeit und funktionierende Anwendbarkeit eines Wissens, über das man selbst nicht Bescheid weiß. Anthony Giddens markiert diese Ähnlichkeit zwischen der menschlichen und der technischen Bereitstellung spezialisierten Wissens und Könnens, indem er beides unter einem Begriff, dem Begriff des Expertensystems, zusammenfasst (vgl. Giddens 1995: 40-41, 49; Schulz-Schaeffer 2000: 210-226).

Oben hatte ich argumentiert, dass die technische Bereitstellung typischer Teilstücke für typische Zwecktätigkeiten einen Einfluss auf die Handlungsziele von Akteur/-innen haben kann. In einen Koffer beispielsweise passt nur eine bestimmte Menge Kleidungsstücke. Wenn man ihn für den Urlaub nutzen möchte, muss man unter Umständen Kleidungsstücke zu Hause lassen, die man gerne mitgenommen hätte. Das mit der Ausdifferenzierung spezialisierter Handlungskompetenzen verbundene Auseinandertreten von intentionalem und regulativem Handlungssinn trägt die Möglichkeit noch sehr viel weitreichenderer Einflussnahmen auf Handlungsziele in sich. Dazu muss man sich nur vergegenwärtigen, dass im zweckrationalen Handeln die Abwägung der Mittel, der Ziele und der Nebenfolgen miteinander in Verbindung stehen. Beispielsweise kann die Behandlung, die den größeren Heilungserfolg verspricht, zugleich auch die für den Patienten/die Patientin belastendere sein oder die vom Routenplaner vorgeschlagene schnellere Zugverbindung diejenige mit dem größeren Risiko, den Anschluss zu verpassen. Über solche Ziel-Mittel-Abwägungen beeinflusst die Handlungsplanung die Spezifizierung der Handlungsziele: Man wollte auf jeden Fall pünktlich zu der Verabredung ankommen, aber muss jetzt nach Fahrtantritt feststellen, dass der Routenplaner anders priorisiert hat. In dem Maße, in dem Akteur/-innen die Handlungsplanung an „Systeme technischer Leistungsfähigkeit oder professioneller Sachkenntnis“ (Giddens 1995: 40) delegieren, delegieren sie damit zwangsläufig auch Entscheidungen an sie, die ihre Handlungsziele betreffen. Dies können – etwa im Fall des stellvertretenden Handelns von Ärzt/-innen oder Rechtsanwäl/-innen – Entscheidungen sein, von denen gilt, dass die Vertretenen „davon in ihrer persönlichen und sozialen ‚Identität‘ und in ihren allgemeinen Existenzmöglichkeiten betroffen sind“ (Weiß 1998: 34).

7. TECHNISCH ERÖFFNETE NEUE HANDLUNGSMÖGLICHKEITEN

Der Übergang zwischen technischen Artefakten, die menschliches Expertentum substituieren, und solchen, die neue Handlungsmöglichkeiten eröffnen, ist empirisch betrachtet fließend. Der Wert der analytischen Unterscheidung zwischen diesen beiden Fällen besteht darin, dass sie auf folgenden Unterschied hinweist: Zwar wird kein auf Geldzählen spezialisierter Kassierer so schnell zählen können wie eine Geldzählmaschine und kaum eine Chauffeurin über das Kartenwissen eines Navigationsgerätes verfügen. Insofern substituieren diese technischen Artefakte nicht nur einfach menschliche Fachleute und Expert/-innen, sondern eröffnen darüber hinaus neue Handlungsmöglichkeiten. Für die handlungstheoretische Analyse entspricht ihre Position im Handlungsgefüge nichtsdestotrotz aber genau der Position des menschlichen Experten, an den die betreffenden Handlungsvollzüge anderenfalls delegiert werden könnten und wurden. Insbesondere haben wir es auch hier in beträchtlichem Umfang mit gesellschaftlich bereits vorgeformten Handlungszusammenhängen zu tun, in denen die Abhängigkeit des Handlungsverlaufs – inklusive der technisch delegierten Handlungsbestandteile – von einem vorgängigen Handlungsziel dementsprechend zumindest in einem gewissen Umfang erhalten bleibt.

In dem Maße dagegen, in dem technische Artefakte neue Handlungsmöglichkeiten eröffnen, wird es auch möglich, ganz neue Handlungsziele und Handlungsentwürfe zu entwickeln, solche also, die noch nicht Bestandteil des gesellschaftlichen Wissensvorrates oder des persönlichen Erfahrungsbestandes sind. Damit fällt – anfangs zumindest – auch die Abhängigkeit der technischen Wirkungen von kulturell verfestigten Handlungszielen und Handlungsmustern weg, die in den beiden zuvor betrachteten Fällen (in unterschiedlichem Umfang) für die sinnhafte Abhängigkeit des technischen Wirkens von vorgängigen Handlungszielen sorgt. Was bedeutet dies für die intentionale Handlungsdimension? Kann man noch sagen, dass der technisch objektivierbare Sinn vom Sinn der Gesamthandlung abhängig ist, für die die Tech-

nik ein Teilstück beisteuert, wenn die Bildung dieser Gesamthandlung von der Existenz dieser Technik abhängt?

Man kann auf diese Frage zwei unterschiedliche Antworten geben: Die klassische handlungstheoretische Antwort lautet, dass menschliches Handeln stets bestimmten ermöglichenden und beschränkenden Bedingungen unterworfen ist, weshalb ein zweckorientiert rationaler Akteur seine Handlungsziele unter Berücksichtigung der jeweils vorliegenden Handlungsmöglichkeiten und -bedingungen bestimmt und seine Handlungen entsprechend plant. Aus dieser Sicht ist der jeweilige Stand der Technik Bestandteil der ermöglichenden und beschränkenden Bedingungen. Ein veränderter Stand der Technik macht diesbezüglich keinen Unterschied. Es gibt deshalb keinen Grund anzuzweifeln, dass die Bedeutung von Technik weiterhin von den Handlungsabsichten der Akteur/-innen abhängig bleibt.

Die gegenteilige Antwort leitet sich aus der Betrachtung der Handlungswirksamkeit technischer Objektivationen ab. Dieser Betrachtung zufolge sind technische Objektivationen anderen kulturell verfestigten Sinnmustern darin ähnlich, dass sie – mit Émile Durkheim gesprochen – als Gussformen für menschliches Handeln wirken (vgl. Durkheim 1984 [1895]: 126), also das durch sie ermöglichte Verhalten begünstigen, wenn nicht gar vorzeichnen. Die erste Antwort macht geltend, dass Handlungsmöglichkeiten stets erst bezogen auf Handlungsziele zu Handlungsmitteln werden. Die zweite Antwort verweist darauf, dass die bereits vorgezeichneten Handlungsoptionen einen beträchtlichen Einfluss darauf haben, welche Handlungsziele Akteur/-innen ausprägen.

Beide Auffassungen sind zweifellos empirisch relevant. Das macht die Frage nach der intentionalen Handlungsdimension zu einer einigermaßen komplizierten Fragestellung. Man betrachte das folgende Beispiel: Unser Wanderer, der eine Schale dabei hat, oder sagen wir besser: eine Flasche, nähert sich einer Quelle. Die in der Gestalt der Flasche sinnhaft angelegte Möglichkeit, sie als Transportmittel für Flüssigkeiten zu benutzen, bringt ihn auf die Idee, sie mit dem Quellwasser zu füllen, um später, falls er durstig werden sollte, etwas zum

Trinken zur Verfügung zu haben. Um des Arguments willen wollen wir annehmen, dass er die Flasche nicht zu diesem Zweck mitführt, sondern aus anderen Gründen; vielleicht, weil sie ein Getränk enthielt, das er bereits ausgetrunken hat. Stammt in diesem Fall die Handlungsintention, die der Handlung zugrunde liegt, von dem Wanderer oder wird sie von der in der Flasche verkörperten Handlungsoption beige-steuert?

Klassisch handlungstheoretisch würde man argumentieren, dass die Handlung auch in diesem Fall an einem menschlichen Handlungsziel orientiert ist und von dorthin ihre Bedeutung erlangt, nämlich an dem Ziel, den zukünftigen Durst zu löschen. Dagegen lässt sich jedoch argumentieren, dass Menschen zu jedem Zeitpunkt ihres Lebens gleichzeitig viele potenzielle Handlungsziele besitzen, ohne dass sie ihnen gleich auch schon so wichtig sind, dass sie Schritte zu ihrer Realisierung unternehmen. Wäre der Wanderer statt an der Quelle an einer Sitzbank vorbeigekommen, hätte er vielleicht einen Augenblick verweilt, mit dem Ziel sich auszuruhen und die frische Luft zu genießen. Wäre er einem Eisverkäufer begegnet, hätte er vielleicht das Bedürfnis verspürt, ein Eis zu essen usw. Aus dieser Sicht sind die in den technischen Artefakten, in der Flasche bzw. der Sitzbank (oder auch in der Dienstleistung des Eisverkäufers) verkörperten Teilstücke von Handlungen dafür verantwortlich, dass ein zu der betreffenden Handlung passendes Handlungsziel von einem potenziellen zu einem aktuellen Handlungsziel des Akteurs wird. Es ist demnach der in den technischen oder sonstigen Objektivationen verkörperte Handlungssinn, der die Akteur/-innen veranlasst, das betreffende Handlungsziel tatsächlich zu verfolgen.

8. ZUSAMMENFASSUNG

In diesem Beitrag habe ich die Frage der relativen Autonomie technischer Artefakte im Rahmen ihrer instrumentellen Nutzung betrachtet. Damit sollte nicht impliziert sein, dass der Einsatz und die Nutzung von Technik in der Arbeitswelt stets und ausschließlich durch instrumentelle Rationalität getrieben sind. Es gibt, wie beispielsweise David Noble (1978) eindrucksvoll gezeigt hat, auch politische Gründe oder – worauf neoinstitutionalistische Studien verweisen – begrenzt rationale Gründe des Nachvollziehens allgemeiner Technologietrends (vgl. z.B. Liao 1996). Nichtsdestotrotz spielt die Betrachtung von Technik als Arbeitsmittel in der Arbeitswelt zweifellos eine zentrale Rolle für den Einsatz und die Nutzung von technischen Artefakten. Dies begründet es, die instrumentelle Nutzung als Bezugsrahmen für die Analyse der relativen Autonomie von Technik zu wählen.

Meine handlungstheoretische Analyse hat allem voran auf die besondere Bedeutung der regulativen Handlungsdimension verwiesen. Die Delegation von Tätigkeiten an technische Artefakte oder Abläufe ist immer eine Delegation in der effektiven *und* in der regulativen Handlungsdimension: Wenn die Herbeiführung einer Wirkung an Technik delegiert wird, geht damit immer auch ein Stück der Steuerung der Handlungsdurchführung an das technische Artefakt über. Genau hierin besteht die ‚Ähnlichkeit‘ technisch realisierter Handlungsbeiträge zu einem großen Teil des menschlichen Handelns. Nicht nur technisches Verhalten ist ein sinnhaft gesteuertes Verhalten auf der Grundlage objektivierten Sinns. Gleiches gilt auch für einen beträchtlichen Teil des menschlichen Handelns, sofern es auf der Grundlage kulturell objektivierten Handlungssinns durchgeführt wird. Wir haben auch gesehen, dass es in der regulativen Handlungsdimension sinnvoll ist, zwischen der Delegation der Steuerung und der Kontrolle zu unterscheiden.

Die handlungstheoretische Analyse hat gezeigt, dass zunehmende Autonomie technischer Artefakte in der regulativen Handlungsdimension auf eine Entkopplung von regulativem und intentionalem Hand-

lungssinn zurückzuführen ist, die ihrerseits eine Begleiterscheinung arbeitsteiliger Spezialisierung ist. In der wissenschaftlich-technischen Zivilisation prägt arbeitsteilige Spezialisierung nicht nur menschliches Zusammenhandeln, sondern zunehmend auch das Zusammenwirken von Mensch und Technik. Zunehmend bekommen wir es nicht nur mit einem menschlichen Expertentum, sondern mit einem in ‚intelligenter‘ Technik verkörperten Wissen und Können zu tun, das dasjenige ihrer Nutzer/-innen so weit übersteigt, dass die technischen Systeme besser wissen als ihre Nutzer/-innen, was es heißt, deren Ziele und Interessen zu verfolgen. Hier beginnt sich nun auch im Bereich des technisch verkörperten Expertentums eine Emanzipation der regulativen Handlungsdimension zu entwickeln – wie zuvor schon bei menschlichem Expertentum, welches die Möglichkeit substanzieller Einflussnahmen auf Handlungsziele in sich trägt.

Der Zusammenhang von regulativer und intentionaler Handlungsdimension in technisch unterstützten Handlungen verändert sich zum anderen auch im Zuge der Ausdifferenzierung von Technikherstellung und -verwendung und des technischen Wandels. Zunehmend werden für alle möglichen kulturell verfestigten Handlungsmuster technische Abläufe bereitgestellt, die als typische Handlungsmittel für typische Handlungsziele fungieren, zunehmend werden sie zu Handlungsmitteln für Ziele, die durch die entsprechenden Techniken überhaupt erst ermöglicht werden. Beides hat zur Konsequenz, dass Akteur/-innen – zumindest: auch – deshalb bestimmte Ziele verfolgen, weil die technischen Mittel dazu verfügbar sind und weil mit ihnen gerade diese Ziele realisierbar sind. Technik beeinflusst hier dann auch in dieser Weise die intentionale Handlungsdimension des betreffenden Handlungsgeschehens.

LITERATUR

- Akrich, Madeline (1992): „The De-Description of Technical Objects“, in: Wiebe E. Bijker/John Law (Hg.), *Shaping Technology; Building Society. Studies in Sociotechnical Change*, Cambridge, Mass. u.a.: MIT Press, S. 205-224.
- Durkheim, Émile (1984 [1895]): *Die Regeln der soziologischen Methode*, hg. und eingeleitet von René König, Frankfurt a.M.: Suhrkamp.
- Freyer, Hans (1966 [1934]): *Theorie des objektiven Geistes. Eine Einleitung in die Kulturphilosophie*, Stuttgart: Teubner.
- Giddens, Anthony (1995): *Konsequenzen der Moderne*, Frankfurt a.M.: Suhrkamp.
- Latour, Bruno (1992): „Where are the Missing Masses? The Sociology of a Few Mundane Artifacts“, in: Wiebe E. Bijker/John Law (Hg.), *Shaping Technology; Building Society. Studies in Sociotechnical Change*, Cambridge, Mass. u.a.: MIT Press, S. 225-258.
- Liao, John (1996): „Information Technology Investment: The Effect of Institutional Isomorphism“, in: *The Journal of High Technology Management Research* 7/1, S. 37-52.
- Luckmann, Thomas (1992): *Theorie des sozialen Handelns*, Berlin u.a.: de Gruyter.
- Luhmann, Niklas (1984): *Soziale Systeme. Grundriß einer allgemeinen Theorie*, Frankfurt a.M.: Suhrkamp.
- Noble, David F. (1978): „Social Choice in Machine Design: The Case of Automatically Controlled Machine Tools, and a Challenge to Labor“, in: *Politics and Society* 8, S. 313-347.
- Parsons, Talcott. (1968 [1937]): *The Structure of Social Action*, 2 Bde., New York: The Free Press.
- Rammert, Werner/Schulz-Schaeffer, Ingo (2002): „Technik und Handeln. Wenn soziales Handeln sich auf menschliches Verhalten und technische Abläufe verteilt“, in: dies. (Hg.), *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*, Frankfurt a.M. u.a.: Campus, S. 11-64.

- Schulz-Schaeffer, Ingo (2000): Sozialtheorie der Technik, Frankfurt a.M. u.a.: Campus.
- Schulz-Schaeffer, Ingo (2007): Zuschriebene Handlungen. Ein Beitrag zur Theorie sozialen Handelns, Weilerswist: Velbrück.
- Schulz-Schaeffer, Ingo (2008a): „Formen und Dimensionen der Verselbständigung“, in: Albert Kündig/Danielle Bütschi (Hg.), Die Verselbständigung des Computers, Zürich: vdf Hochschulverlag, S. 29-53.
- Schulz-Schaeffer, Ingo (2008b): „Technik“, in: Nina Baur/Hermann Korte/Martina Löw/Markus Schroer (Hg.), Handbuch Soziologie, Wiesbaden: VS Verlag, S. 445-463.
- Schulz-Schaeffer, Ingo (2008c): „Technik als sozialer Akteur und als soziale Institution. Sozialität von Technik statt Postsozialität“, in: Karl-Siebert Rehberg (Hg.), Die Natur der Gesellschaft, Verhandlungen des 33. Kongresses der Deutschen Gesellschaft für Soziologie in Kassel 2006, Frankfurt a.M. u.a.: Campus, S. 705-719.
- Schütz, Alfred/Luckmann, Thomas (1984): Strukturen der Lebenswelt, Bd. 2, Frankfurt a.M.: Suhrkamp.
- Wagner, Gerald (1994): „Vertrauen in Technik“, in: Zeitschrift für Soziologie 23/2, S. 145-157.
- Weber, Max (1972 [1922]): Wirtschaft und Gesellschaft. Grundriß der verstehenden Soziologie, besorgt von Johannes Winckelmann, Tübingen: Mohr.
- Weiß, Johannes (1998): Handeln und handeln lassen. Über Stellvertretung, Wiesbaden: Westdeutscher Verlag.
- Wiebe E. Bijker/John Law (Hg.), Shaping Technology; Building Society. Studies in Sociotechnical Change, Cambridge, Mass. u.a.: MIT Press.

Herausforderungen der sozio-technischen Evaluation der Arbeit mit Autonomen Systemen

Thomas Herrmann und Jan Nierhoff

1. EINLEITUNG UND MOTIVATION EINES HEURISTIK-BASIERTEN ANSATZES

Arbeitsbedingungen, insbesondere im industriellen Kontext, sind maßgeblich durch die eingesetzten Technologien bestimmt. Wir erachten die Evaluation von Arbeitsbedingungen im Kontext Autonomer Systeme aus zwei Gründen als herausfordernd:

(1) Menschliche Akteur/-innen in vielfältigen Rollen interagieren mit Technologie, die durch autonome, dezentralisierte cyberphysikalische Systeme gekennzeichnet ist. Die daraus entstehende reziproke Beeinflussung zwischen Mensch und Technik konstituiert ein komplexes sozio-technisches Szenario (vgl. Baxter/Sommerville 2011: 4-17). Es ist gekennzeichnet durch die Verzahnung technischer Komponenten mit organisationalen Maßnahmen zur Unterstützung von Kommunikation, Kollaboration und Koordination. Sozio-technische Systeme können nur unvollständig beschrieben bzw. dokumentiert werden (vgl. Suchman 1995: 56-64) und unterliegen ständiger Weiterentwicklung (vgl. Fischer/Herrmann 2011: 1-33).

(2) Autonome Systeme bedürfen einer kontinuierlichen Anpassung auf wechselnde Herausforderungen und zur Wahrnehmung von Optimierungspotenzialen. Diese Anpassung ist zum einen durch menschliche Akteur/-innen zu leisten und kann zum anderen durch maschinelles Lernen unterstützt werden, falls ein entsprechender Entwicklungsstand des Autonomen Systems vorliegt. Daraus resultiert eine besondere Dynamik.

Wir betrachten ein technisches System als umso autonomer, je eher es seine Leistung und seinen Nutzen ohne die Mitwirkung eines menschlichen Akteurs/einer menschlichen Akteurin erbringen kann (vgl. Schmidt/Herrmann 2017). Das bedeutet, dass keine feingranulare Interaktion und keine anhaltende Aufmerksamkeit im Verlauf der Leistungserbringung durch das Autonome System erforderlich sind. Im Unterschied zu einfachen Automaten (etwa einer herkömmlichen Kaffee- oder einer Waschmaschine) sind neuere Entwicklungen auf dem Gebiet Autonomer Systeme zum einen durch eine ständige Steuerung mittels Kontextauswertung gekennzeichnet, die den menschlichen Input kompensiert. Man spricht auch von impliziter Interaktion; hierfür ist eine geeignete Sensorik notwendig. Zum anderen passen neuere Autonome Systeme die ihnen initial vorgegeben Regeln als Teil ihres Leistungsspektrums selbst an bzw. unterstützen diese Anpassung auf Basis der Historie der Auswertung des Kontextes und des gelegentlichen menschlichen Systeminputs.

Die Aufgaben, die im Hinblick auf diese Anpassungsvorgänge und gelegentlichen Eingriffe beim Menschen verbleiben, sind keineswegs marginal oder beiläufig, wie unten (Kapitel 5) zu zeigen sein wird. Um der Komplexität und der Dynamik dieser Aufgaben beim sozio-technischen Design zu begegnen, schlagen wir den Einsatz von Heuristiken zur agilen Systemevaluation vor. Während der Heuristik-Begriff Domänen-übergreifend sehr allgemein als „Lehre, Wissenschaft von dem Verfahren, Probleme zu lösen [altgriechisch εὐρίσκειν (heuriskein): entdecken, finden, d.V.]“ definiert ist (Brockhaus Enzyklopädie 1995: 1568), zeichnen sich als heuristisch benannte Vorgehensweisen meist dadurch aus, dass mit begrenztem Wissen eine ausreichend gute Lö-

sung gefunden werden soll. Im psychologischen Kontext definiert Zimbardo (2004: 371) Heuristiken als „kognitive Eilverfahren, die bei der Reduzierung des Bereichs möglicher Antworten oder Problemlösungen nützlich sind, indem sie Faustregeln als Strategien anwenden“. Im Gegensatz zu Algorithmen garantieren Heuristiken keine Lösungen (vgl. Wessells 1982) und sind fehleranfälliger (vgl. Stangl 2018). Während heuristische Ansätze also keine ‚100-Prozent-Lösungen‘ versprechen, bieten sie eine durch Pragmatik geprägte Vorgehensweise, um die kritischsten Probleme mithilfe eines überschaubaren Aufwands zu identifizieren. Das wohl bekannteste Beispiel eines solchen Ansatzes im Bereich der Mensch-Technik-Interaktion sind Nielsens Heuristiken zur Usability Evaluation von interaktiven Systemen (vgl. Nielsen 1994: 152-158). Im Rahmen des Forschungsprojekts *Heuristiken für die Industrie 4.0*¹ wurde ein Heuristik-Set zur Evaluation sozio-technischer Systeme entwickelt.

Abbildung 1 zeigt, wie ein solches Heuristik-Set zur Unterstützung einer sozio-technischen Systemevaluation eingesetzt werden kann: Eine zentrale Rolle spielen die Personen, die sie durchführen. Ihre Erfahrung mit dem zu analysierenden System, ihre Expertise in einer spezifischen Domäne und auch ihre Vertrautheit mit dem Einsatz der Heuristiken haben Einfluss auf das Ergebnis der Evaluation. Zielgruppen der Heuristiken sind sowohl Planer/-innen, Designer/-innen, Manager/-innen und Ingenieur/-innen, die im Kontext von Industrie-4.0-Lösungen agieren, als auch die dabei tätigen operativen Kräfte, ihre Interessenvertretung sowie Berater/-innen und Ausbilder/-innen. Im Unterschied zu abstrakten Gestaltungsgrundsätzen (vgl. Grote 2015), haben Heuristiken den Anspruch, dass sie auch für die Betroffenen selbst anwendbar sind. Sie sind damit nicht in erster Linie für die Nutzung durch Expert/-innen für Arbeitsgestaltung und sozio-technisches De-

1 Gefördert durch das Ministerium für Kultur und Wissenschaft des Landes NRW im Rahmen des Themenbereichs *Digitalisierung von Arbeit – Industrie 4.0* des Forschungsinstituts für gesellschaftliche Weiterentwicklung e.V.

sign vorgesehen, sondern für diejenigen, die aus anderer Perspektive mit der Entwicklung und der Einführung eines Industrie-4.0-Konzeptes befasst sind.

Abbildung 1: Anwendungskonzept der Heuristiken zur sozio-technischen Systemevaluation



Quelle: eigene Darstellung

Die Analyse des Systems kann auf verschiedenen Grundlagen aufbauen, etwa auf Beobachtungen eines konkreten Settings (z.B. einer Fertigungsanlage), auf Basis von Modellen der Arbeitsprozesse oder indirekt auf Interviews mit Nutzer/-innen oder mit Planer/-innen des Systems. Die Heuristiken dienen als Impulse, das System unter Berücksichtigung der von ihnen adressierten Aspekte zu betrachten. Sie helfen, im Zusammenspiel mit dem eigenen Expertenwissen, kritische Aspekte des sozio-technischen Systemdesigns zu fokussieren. Die Konkretisierung, die durch die Verschriftlichung eher abstrakter Konzepte, wie z.B. Flexibilität, zu Heuristiken erfolgt, macht diesbezügliche Probleme im System benennbar und unterstützt die Kommunikation zwischen den verschiedenen Akteur/-innen. So werden potenziell übersehene Defizite und Verbesserungsanforderungen aufgezeigt und behandelbar gemacht. Erhebungen am Beispiel der Digitalisierung von

Industrieanlagen bzw. der dafür entwickelten Konzepte machen deutlich, dass über eine reine Problemidentifikation hinaus mithilfe der Heuristiken auch die Erhebung und das Verständnis der Lösungsansätze aus einer sozio-technischen Perspektive (vgl. Herrmann 2012) unterstützt wird.

Der vorliegende Beitrag stellt einen ersten Kandidaten für ein solches Heuristik-Set vor und gibt Hinweise zu dessen Einsatzmöglichkeiten. Die Vorgehensweise bei der Entwicklung der Heuristiken wird offengelegt und verdeutlicht, warum deren Anwendbarkeit auf sozio-technische Konzepte im Allgemeinen auch auf die Industrie-4.0-Konzepte unter Einbeziehung Autonomer Systeme zutrifft.

2. IDENTIFIKATION EINES INITIALEN HEURISTIK-SETS

Wir verstehen die Integration Autonomer Systeme in Arbeitsvorgänge als ein sozio-technisches System. Um die relevanten Aspekte für ein solches sozio-technisches Design möglichst umfassend zusammenzustellen, wurde Literatur aus den Bereichen Human-Computer Interaction, Computer Supported Cooperative Work, Process Redesign, Socio-Technical Design, Job-Design und Privacy gesichtet. Extrahiert wurden 174 einzelne Inhaltsaspekte, die auf gelungene Systemgestaltung und auf potenzielle Probleme hinweisen. Tabelle 1 listet die gesichtete Literatur und die Anzahl der extrahierten Inhaltsaspekte für die Heuristiken.²


2 Streng genommen handelt es sich bei den in der Literatur gefundenen Items bereits um Heuristiken bzw. Richtlinien oder Prinzipien. Um Konfusion zu vermeiden, benutzen wir aber den Begriff *Inhaltsaspekte*. Es werden also diese *Inhaltsaspekte* zu neuen *Heuristiken* gruppiert. Die Gesamtmenge der neuen sozio-technischen *Heuristiken* ist ein *Heuristik-Set*.

Tabelle 1: Übersicht der gesichteten Literatur und extrahierter Items

Domäne	Autor/-innen	Anzahl extrahierter Inhaltsaspekte
Socio-Technical Design	Clegg 2000	15
	Cherns 1987	11
	Eason 1988	10
Human-Computer Interaction	Nielsen 1994; Nielsen/Molich 1990	11
	Shneiderman et al. 2010	9
	Dix et al. 2004	14
	Schneider 2008; referenziert ISO 9241-110: 2006	7
Computer-Supported Cooperative Work	Greenberg et al. 1999	5
	Baker/Greenberg/ Gutwin 2001	7
	Herrmann/Wulf/Hartmann 1996	8
Socio-Technical Design; Job Re-Design	Mumford 1983	5
Job Re-Design	Hackman/Oldham 1975	7
	Dunckel 1989	10
	Grote et al. 2000; Grote 2015	17
Privacy	Rost/Bock 2011	6
	Clement 1993	5
Process Re-Design	Reijers/Mansar 2005	27
Summe		174


Exemplarisch zeigt die folgende Abbildung Untermengen der gefundenen Inhaltsaspekte aus der Literatur, die zu den Heuristiken (i.e. Clustern) *Autonomie* (auch Themen wie Flexibilität abdeckend) und *Evolutionäre Weiterentwicklung* (Adaption) verdichtet wurden.

Abbildung 2: Auszug von Inhaltsaspekten, die zu den neuen Heuristiken Autonomie und Evolutionäre Weiterentwicklung gruppiert wurden



Autonomie

(Grote 2015) Kontrollierbarkeit der Technik durch Menschen
 (Dix et al. 2004) Flexibility. Multithreading (ability of system to support user interaction for several tasks at a time, concurrent multimodality: simultaneous communication of information pertaining to separate tasks, interleaving multimodality: permits temporal overlap between separate tasks, dialog is restricted to a single task).
 (Nielsen 1994) Flexibility and efficiency of use. Accelerators – unseen by the novice user – may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.
 (Rost/Bock 2011) Ability to intervene (contingency) – to operationalise especially data subject rights and the ability of information processing entities respective operators of systems to demonstrate verifiable that they actually have steering control over their systems and are not dominated by the system.
 (Hackman/Oldham 1975) Autonomy. The degree to which the job provides substantial freedom, independence, and discretion to the employee in scheduling the work and in determining the procedures to be used in carrying it out.
 (Dunckel 1989) Zeitspielraum. Möglichkeiten für die Arbeitenden den Handlungsablauf selbständig zeitlich zu strukturieren; zu enge zeitliche Vorgaben können einen (ansonsten) recht großen Handlungsspielraum wieder zunichtemachen.



Evolutionäre Weiterentwicklung

(Eason 1988) To serve the functional needs of the organization by serving the functional needs of individual users a major form of organizational and individual learning is required. A progressive, planned form of evolutionary growth complements existing design procedures and organization change practices.
 (Clegg 2000) Systems and their design should be owned by their managers and users. This amends Cherm's principle of compatibility and involves a change from his emphasis on user participation to user ownership.
 (Eason 1988) The effective exploitation of socio-technical systems depends upon adoption of a planned process of change that meets the needs of people.
 (Herrmann/Wulf/Hartmann 1996) Group-oriented configurability. Enables each group of users to specifically select the appropriate number of functions and their functional alternatives during a process of participative configuration.
 (Cherns 1987) Transitional organization. See the design team and its process as a vehicle of transition.
 (Cherns 1987) Incompletion or the Forth Bridge Principle. We all know that the present period of transition is not between past and a future stable state but really between one period of transition and another.

Quelle: eigene Darstellung

Diese beiden Aspekte sind von besonderer Bedeutung im Kontext Autonomer Systeme, weil sich bei deren Einsatz die Frage ergibt, wie die Autonomie derjenigen beeinflusst wird, die die Systeme vorbereiten, nutzen und warten. Die Anpassungsfähigkeit Autonomer Systeme – etwa auf dem Weg des maschinellen Lernens, impliziert eine stetige Weiterentwicklung, von der menschliche Akteur/-innen betroffen sind und für die sich die Frage stellt, inwieweit sie von diesen Akteur/-innen beeinflussbar ist.

2.1 Vorgehensweise beim Clustering und methodische Herausforderungen

Die Verdichtung der 174 Inhaltsaspekte zu neuen Heuristiken wurde in drei Runden von fünf Expert/-innen aus den Bereichen Informatik, Soziologie, Arbeitswissenschaften und Prozessmanagement durchgeführt. Da bei der Menge und Art (semantische Vagheit, zum Teil wenig Trennschärfe) der Inhaltsaspekte kein eindeutiges, ‚korrektes‘ Gruppierungsschema vorgegeben werden konnte, wurde ein erster Gruppierungsvorschlag unterbreitet, auf den in mehreren Iterationen Zuordnungsversuche durch die Expert/-innen aufbauten sowie eine anschließende Diskussion der Problemfälle. So konnten Diskrepanzen beim Verständnis der Inhaltsaspekte und der Beziehungen unter ihnen schrittweise aufgeklärt werden.

Im Verlauf der Zuordnungen und ihrer Erörterung wurde deutlich, dass die spätere Formulierung der Heuristiken eine weitere Herausforderung darstellt. Die Beschreibung einer Heuristik muss einerseits einen hohen Abstraktionsgrad haben, um den vielfältigen Inhaltsaspekten, die sie umfasst, gerecht zu werden. Andererseits muss sie konkret genug sein, um leicht verständlich und gut merkbar zu bleiben, damit die unterstellte Wirkungsweise der Heuristik (siehe Kapitel 4) zur Geltung kommen kann. Als Vorbereitung für den nächsten Arbeitsschritt – die formative Evaluation des Heuristik-Sets – wurden die Heuristiken zur Erleichterung des Verständnisses jeweils in *Sub-Heuristiken* untergliedert, welche spezifische Aspekte der Heuristik fokussieren.

2.2 Ergebnis der Literatur-basierten Identifikation von Heuristiken

Abbildung 3: Auf Basis der Literaturrecherche identifiziertes Heuristik-Set



Quelle: eigene Darstellung

Das Ergebnis der Literaturrecherche ist das in Abbildung 3 dargestellte Set von 13 Heuristiken (A-M). Zur Erhöhung ihrer Handhabbarkeit (z.B. wenn sie zur Strukturierung eines Interviews genutzt werden) wurden die Heuristiken in vier Themenfeldern organisiert.³

Die Sub-Heuristiken zweier Heuristiken sind im Folgenden exemplarisch gelistet:

Heuristik E: Unterstützung von Autonomie und Flexibilität

E1: Minimierung von Auflagen, die Nutzer/-innen von frei, flexibel und individuell getroffenen Entscheidungen abhalten könnten (zu Arbeitsabläufen, einzusetzenden Mitteln und Vorgehensweisen, zu nutzenden Informationsräumen, Zeitmanagement etc.)

E2: Autonomie und Selbst-Regulierung in Bezug auf vielfältige Weisen der Aufgabenausführung sowie die Arbeitsaufteilung zwischen Mensch und Maschine und zu beherrschende Schwierigkeitsgrade.

E3: Nutzer/-innen haben ein Empfinden von Kontrolle; sie können ungewollte Belastung und Stress vermeiden, die Aufdeckung personenbezogener Daten verhindern und werden befähigt, bei unerwarteten, sie betreffenden Prozessen zu intervenieren.

E4: Die Koordination zwischen Teilnehmer/-innen kann flexibel verhandelt und durch sie selbst gesteuert werden, z.B. innerhalb autonomer Gruppen.

Heuristik F: Unterstützung von Anpassung, Wandel und evolutionärer Weiterentwicklung

F1: Unterstützung der kontinuierlichen Weiterentwicklung und der Anpassung des Systems durch alle Teilnehmer/-innen (insbesondere Nutzer/-innen) und Stakeholder. Systemische Interdependenzen, Unvollständigkeit und Kontingenz werden als inhärente Charakteristiken

3 Diese Zuordnung geschah intuitiv und wurde in neueren Iterationen des Heuristik-Sets aufgehoben.

des sozio-technischen Systems anerkannt und erfordern kontinuierliche Evolution.

F2: Kontinuierlicher Wandel wird auf individuellem und auf organisationalem oder Gruppen-Level unterstützt. Die den Wandel formenden Entscheidungen, die die Kollaboration beeinflussen, müssen verhandelbar sein.

F3: Möglichkeiten des Wandels und der Evolution müssen von Anfang an geplant werden und ein inhärenter Bestandteil des sozio-technischen Systems sein. Die Zuordnung von Aufgaben muss die individuellen Unterschiede der Teilnehmer/-innen berücksichtigen.

F4: Vorgegebene Workflows müssen adaptierbar sein.

3. FORMATIVE EVALUATION DES HEURISTIK-SETS MITHILFE EINER DATENBANK SOZIO-TECHNISCHER PROBLEMBERICHTE

Analog zur Vorgehensweise von Nielsen (vgl. 1994: 152-158) soll die Tauglichkeit des Heuristik-Sets mithilfe von Problemen, die bei tatsächlich implementierten sozio-technischen Lösungen erkannt wurden, geprüft werden. Dabei geht es vorrangig um drei Fragen:

- Können alle Probleme einer oder mehreren Heuristiken zugeordnet werden und kann es somit als wahrscheinlich gelten, dass jedes Problem erkannt wird?
- Lässt sich ein Ranking der Heuristiken erkennen, insbesondere im Hinblick auf die Zahl der Probleme, denen sie zugeordnet werden können?
- Welche Hinweise auf die Erweiterung der Heuristiken in Verbindung mit den Zuordnungsversuchen gibt es?

Die Beantwortung dieser Fragen zielt auch praktisch darauf, die Menge der Heuristiken neu zu strukturieren, um Überlappungen zu vermeiden, weniger wichtige Aspekte auszusortieren etc.

3.1 Aufbau der Datenbank sozio-technischer Problembereiche

Zur Untersuchung der Fragen wurde eine Datenbank aufgebaut, die zum Zeitpunkt der ersten Testung des Heuristik-Sets (März 2018) 223 Probleme aus den folgenden neun allgemeinen sozio-technischen Fällen (vgl. Tabelle 2) enthält:

Tabelle 2: Beschreibung der neun sozio-technischen Fälle

#	Fall
1	Koordination zwischen Zahnmedizin-Studierenden und Ausbilder/-innen während des Praktikums mittels Datenbrillen
2	Bestellung und Koordination von Dienstleistungen für ältere Menschen (vgl. Herrmann/Prilla/Nolte 2016)
3	KreativBarometer – kontinuierliche Erfassung relevanter Arbeitsbedingungen für das Kreativitätsklima am Arbeitsplatz (vgl. Nierhoff/Herrmann 2017)
4	Elektronische Systeme im Gesundheitswesen (Workshop zu zwölf Studien) (vgl. Herrmann et al. 2017)
5	Unterstützung der Reflexion von Gesprächen mit Angehörigen von Schlaganfallpatienten (vgl. Prilla/Herrmann 2017)
6	Elektronische Nutzung räumlich verteilter Laborexperimente für Ingenieur-Studierende
7	Aufbau einer Orientierungseinheit für Erstsemester mittels Augmented Reality
8	Kontinuierliche Aktualisierung von Webseiten kleinerer Organisationseinheiten
9	Elevated – strategische Verbesserungsplanung einer Schule anhand von Daten

Typische Einträge in der Problem Datenbank sind etwa:

- Dem Lehrer wurde es ermöglicht, auf Schüler-Anfragen mit Text-basierten Nachrichten zu antworten; dies wurde im Vergleich zu den favorisierten Sprachnachrichten als unangenehm wahrgenommen. (Fall 9)
- Die Patienten – welche nur indirekt in das Training der Zahnmedi-zin-Studierenden involviert sind – mussten verstehen was passiert: Warum wurden Daten gesammelt oder Nachrichten geschrieben? (Fall 1)

Diese Fälle waren aus der Zeit von 2010 bis 2017 dokumentiert. Paral-lel zur ersten Testung wurde die Problem Datenbank um Fälle erweitert, die sich speziell mit Industrie-relevanten Konzepten und dem Einsatz Autonomer Systeme befassen. Dies sind im Einzelnen:

- Self-Learning Manufacturing Workplace – digitale Lösungen in der Produktion von Komponenten für automatisierte Fertigungsstraßen; beinhaltet Wartungs-Koordination, Visualisierung unterstützender Daten und Trend-Analysen zu Defekten.
- Produktion von Zahnimplantatsteilen – Erzeugung von Individuallö-sungen, aber auch Serienproduktion im Bereich der Dentaltechnik.
- Predictive Maintenance – vorausschauende Wartung in der Karosserie-montage eines Automobilherstellers.
- Assistenzsysteme – Mitarbeiterunterstützung bei der Ausführung von Fertigungsaufgaben durch die (unterschiedlich strikte) Vermitt-lung von Anleitungen zur Vorgehensweise.

Diese Erweiterung dient der Überprüfung der Frage, ob die Ergebnisse zu der Anwendbarkeit der Heuristiken, die aus der Betrachtung der all-gemeinen, eher aus dem Dienstleistungsbereich stammenden Fälle entwickelt werden konnten, sich auch auf den Industriebereich übertra-gen lassen.

3.2 Softwaregestütztes Zuordnungsexperiment zur Bewertung des initialen Heuristik-Sets

In einem von sechs Expert/-innen durchgeführten Experiment wurden 223 Einträge aus der Problemdatenbank den (Sub-)Heuristiken zugeordnet. Bei diesem Experiment wurde jeweils eingeschätzt,

- wie gut das Verständnis eines Problems ist,
- wie schwerwiegend es ist,
- welche Heuristik sich am besten zuordnen lässt,
- wie sicher man sich bei dieser Zuordnung fühlt,
- welche Sub-Heuristiken, die zu der am besten passenden gehören, sich dem Problem zuordnen lassen,
- welche weiteren Heuristiken eventuell zu dem Problem passen können.

Sollte keine Heuristik als passend erscheinen, gab es die Möglichkeit, die Option ‚Zusätzliche Heuristik erforderlich‘ zu markieren; auch die Sicherheit bei dieser Entscheidung wird durch die Proband/-innen bewertet. Darüber hinaus hatten die Teilnehmer/-innen die Möglichkeit, zu allen Skalenabfragen Kommentare einzugeben.

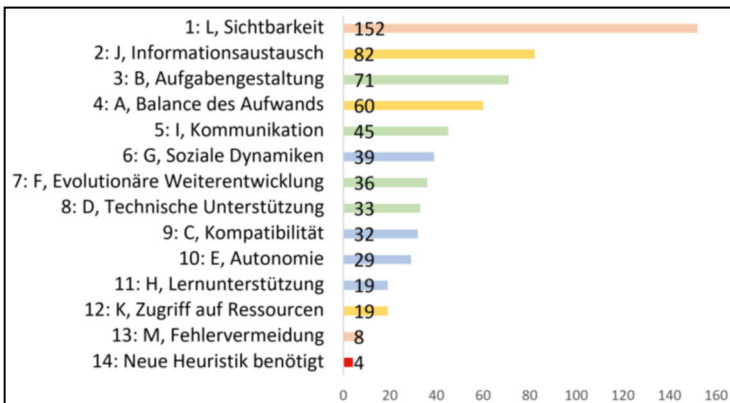
3.3 Ergebnisse der Zuordnung zwischen Problemdatenbank und Heuristik-Set

Von 223 Problemen wurde nur in einem Fall von vier Expert/-innen angegeben, dass hier eine zusätzliche Heuristik benötigt würde. Bei vier weiteren Problemen, waren jeweils zwei Expert/-innen dieser Ansicht. Bei weiteren 27 Problemen meinte jeweils ein Experte/eine Expertin, dass hier eine zusätzliche Heuristik benötigt wird.

Die weitere Auswertung betrachtet nur diejenigen Probleme (insgesamt 217), für die höchstens ein Experte/eine Expertin angibt, dass eine zusätzliche Heuristik notwendig sei. Bei den Sub-Heuristiken wurde bei 1718 Auswahlen 74-mal gefordert, dass eine neue Sub-Heuristik

eingeführt werden sollte. Weiterhin werden diejenigen Probleme nicht weiter betrachtet (insgesamt 18), bei denen die durchschnittliche Verständnissicherheit hinsichtlich des Problems und die Zuordnungssicherheit zu den Heuristiken jeweils nicht größer ist als vier (auf einer Skala von eins bis sieben; sieben entspricht der höchsten Sicherheit). Daraus kann gefolgert werden, dass 199 von 223 Problemen (das entspricht 89 Prozent) mit relativer Sicherheit mit einer oder mehreren der getesteten Heuristiken identifizierbar sind.

Abbildung 4: Ranking der Zuordnungshäufigkeit der Probleme zu Heuristiken⁴



Quelle: eigene Darstellung

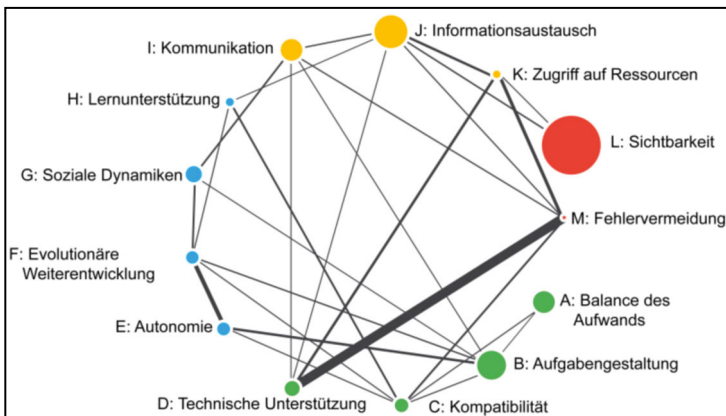
Um zu beurteilen, welche Heuristiken in welchem Umfang zugeordnet wurden, konzentrieren wir uns auf 106 (der 199) Probleme, bei denen sich die Expert/-innen hinsichtlich der Zuordnung relativ einig sind. Anhand von Abbildung 4 wird deutlich, dass die Verwendung der Heuristiken sehr unterschiedlich verteilt ist. Die sechs Expert/-innen haben

4 Berücksichtigt wurden die Zuordnungen zu 106 Problemen mit der relativ höchsten Zuordnungssicherheit durch sechs Expert/-innen.

für jedes der 106 Probleme jeweils einen Punkt auf eine oder mehrere Heuristiken verteilt, so dass insgesamt 636 Punkte vergeben wurden. 152 Punkte entfallen dabei auf *Sichtbarkeit* und nur acht Punkte auf *Fehlervermeidung*. Abbildung 4 zeigt eine Rangverteilung; die ersten vier und die letzten drei Plätze sind weitgehend unabhängig davon, wie viele Probleme man in Abhängigkeit von der Zuordnungssicherheit und der Homogenität zwischen den Expert/-innen auswählt, während die mittleren Plätze ihre Position dabei verändern.

Aufgrund der ungleichen Verteilung der Heuristikzuordnungen bietet es sich an, die Heuristiken zu bündeln, um eine etwas gleichmäßigere Verteilung in Verbindung mit der Reduktion der Anzahl der Heuristiken zu erreichen. Dazu wurde überprüft, welche Heuristiken häufig gemeinsam denselben Problemen zugeordnet werden. Daraus ergibt sich das in Abbildung 5 gezeigte Zusammenhangsdiagramm.

Abbildung 5: Zusammenhangsdiagramm der Heuristiken⁵



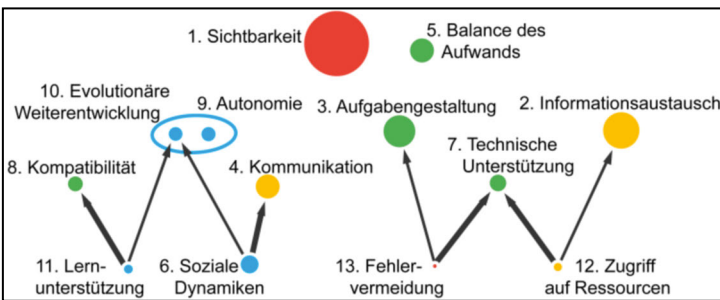
Quelle: eigene Darstellung

⁵ Je stärker die Verbindungslinie, umso häufiger wurden Probleme beiden entsprechenden Heuristiken zugeordnet. Dargestellt sind die 27 stärksten Zusammenhänge (von 78 möglichen insgesamt).

Aufgrund der gezeigten Zusammenhänge bietet es sich an, die inhaltlichen Aspekte der Heuristiken mit geringerer Zuordnungszahl zusammenzufassen bzw. anderen Heuristiken zuzuordnen. Ein solcher Zuordnungsvorschlag wird in Abbildung 6 dargestellt.

Dabei ist darauf zu achten, dass die Integration einer Heuristik in eine andere nicht immer vollständig gelingen kann. So kann *Fehlervermeidung* nur zum Teil im Zusammenhang mit *Technische Unterstützung* angesprochen werden, da sie auch im Zusammenhang mit Aufgabengestaltung relevant ist.

Abbildung 6: Reorganisation der Heuristiken basierend auf dem Zuordnungsexperiment



Quelle: eigene Darstellung

Basierend auf den erkannten Zusammenhängen wurden die 13 untersuchten Heuristiken zu acht neu formulierten Heuristiken zusammengefasst (vgl. Abbildung 7).

Dabei fällt im Kontext dieses Beitrages insbesondere auf, dass *Autonomie* und *Evolutionäre Weiterentwicklung* sich als Kandidaten für eine Zusammenfassung eignen (in der Neufassung Nr. 2, *Flexibilität*). Zum einen eröffnet die mit Autonomie verbundene Variierung von Vorgehensweisen sowie der flexible Einsatz von Technik Spielräume, um Veränderungen zu erproben. So bilden sich Kompetenzen und Erfahrungen, um sich an der Weiterentwicklung des gesamten sozio-technischen Systems zu beteiligen. Wie oben ausgeführt ist eine solche

kontinuierliche Anpassung und Weiterentwicklung beim Einsatz Autonomer Systeme relevant. Sie sind durch Kontextauswertung und implizite Anpassung ihres Verhaltens an sich ändernde Umweltbedingungen geprägt. Zum anderen setzt eine stetige Weiterentwicklung autonomer technischer Systeme im sozio-technischen Gesamtsystem voraus, dass die betroffenen menschlichen Akteur/-innen über ausreichend Handlungsspielraum verfügen, um mit den Anpassungen umzugehen oder sie auch proaktiv zu veranlassen.

Abbildung 7: Neu zusammengefasstes Heuristik-Set⁶



Quelle: eigene Darstellung

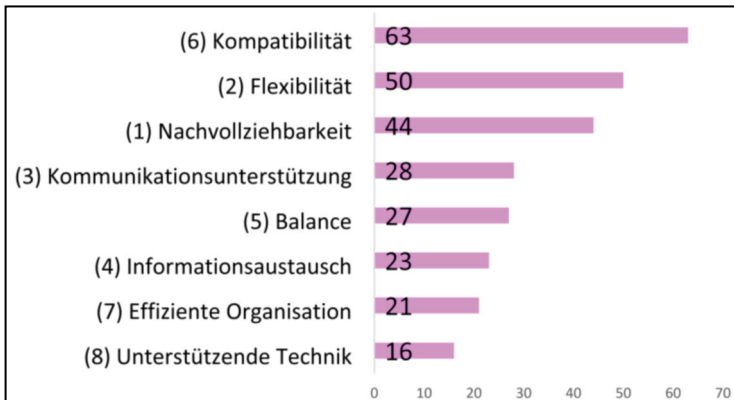
⁶ Die Hervorhebungen kennzeichnen das Stichwort, unter dem auf die jeweilige neue Heuristik im Folgenden Bezug genommen wird.

3.4 Passung der Heuristiken auf Industriekonstellationen mit Autonomen Systemen

Die acht zusammengefassten Heuristiken wurden in einem weiteren Zuordnungsexperiment auf die oben genannten Industriefälle angewendet. Dazu wurden fünf Expert/-innen gebeten, für 82 Probleme anzugeben, inwieweit jede der acht Heuristiken jeweils passend ist (Wert 7) oder nicht (Wert 1). Es zeigt sich, dass die Heuristiken auch in diesem Kontext anwendbar sind: Bei allen Problemen haben jeweils mindestens zwei Personen eine Heuristik mit dem Wert 6 oder 7 zugeordnet. Allerdings werden nicht immer die gleichen Heuristiken für ein Problem ausgewählt: In den elf heterogensten Fällen wurden fünf oder sechs verschiedene Heuristiken demselben Problem zugeordnet; insgesamt 16 Fällen ist keine eindeutige Priorität für eine Heuristik erkennbar. In 66 Fällen ist jedoch eine deutliche Priorisierung der zugeordneten Heuristiken erkennbar.

Auf der Basis dieser 66 Fälle ergibt sich eine Rangfolge, bei der *Kompatibilität* auf dem ersten, *Flexibilität* auf dem zweiten und *Nachvollziehbarkeit* auf dem dritten Platz stehen. Diese Rangfolge differiert deutlich zu der ursprünglichen Rangfolge (vgl. Abbildung 2), bei der *Kompatibilität* und die zu *Flexibilität* zusammengefasste *Autonomie* und *evolutionäre Entwicklung* deutlich schwächer abschneiden. Diese quantitative Veränderung der Rangfolge korrespondiert mit dem inhaltlichen Tatbestand, dass es bei Autonomen Systemen im industriellen Kontext zum einen eine besondere Herausforderung ist, ob deren Verhalten mit den Anforderungen und den Erwartungen des Umfeldes kompatibel ist. Zum anderen ergibt sich ein Spannungsfeld zwischen der Autonomie technischer Systeme und der Autonomie (Flexibilität) der betroffenen Mitarbeiter/-innen, wobei auch die Frage nach dem verbleibenden Spielraum für eine Anpassung der Technik relevant ist. Insgesamt sind auch die anderen Heuristiken relevant, wie Abbildung 8 zeigt; die Unterschiede hinsichtlich der Zuordnungshäufigkeit sind deutlich geringer geworden im Vergleich zu dem ersten Zuordnungsexperiment.

Abbildung 8: Ranking der Zuordnungshäufigkeit der Probleme zu Heuristiken⁷



Quelle: eigene Darstellung

4. ANWENDUNG DER HEURISTIKEN

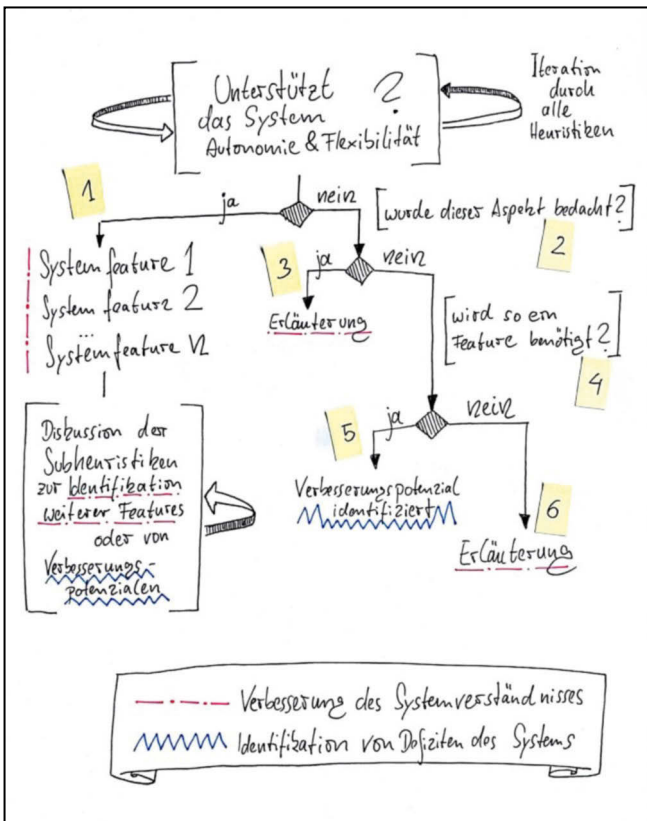
Dieser Abschnitt zeigt zwei Einsatzmöglichkeiten der Heuristiken auf. Während der Ablauf einer auf Heuristiken basierenden Systemexploration in Kapitel 4.1 schematisch beschrieben wird, wird in Kapitel 4.2 der konkrete Einsatz der Heuristiken zum Zwecke einer Systemanalyse erläutert. In beiden Vorgehensweisen werden die Analysen von mit den Heuristiken vertrauten Expert/-innen durchgeführt; es sei angemerkt, dass dies eine sinnvolle Vorgehensweise ist, die Heuristiken aber den Anspruch haben, auch durch ‚Heuristik-Laien‘, wie z.B. in dem System tätige Personen, verstehbar und handhabbar zu sein (vgl. Kapitel 1).

7 Ranking von fünf Expert/-innen; Basis: Von 82 Problemfällen wurden nur die 66 Fälle mit der relativ höchsten Zuordnungssicherheit berücksichtigt.

4.1 Heuristik-basierte Systemexploration

Beispielhaft sei ein Vorgehensmodell für die Exploration eines Systems auf Basis eines durch die Heuristiken semi-strukturierten Interviews mit einem Systemdesigner erläutert (siehe Abbildung 9):

Abbildung 9: Exploration eines Systemdesigns mithilfe der Heuristiken



Quelle: nach Schafler et al. 2018, Übersetzung d.V.

Der Prozess iteriert durch die Heuristiken und überprüft für jede, ob das System entsprechende Eigenschaften aufweist oder nicht und ob letzteres ein Problem ist (z.B. „Unterstützt das System Autonomie & Flexibilität?“). Ist dies der Fall (1), können die Systemfeatures, die diese Heuristik unterstützen, ausführlicher erkundet werden. Eine Diskussion der zugehörigen Sub-Heuristiken kann weitere Features oder Verbesserungspotenziale aufdecken. Wird eine Heuristik nicht durch das System behandelt (2), muss überprüft werden, ob dies absichtlich so design wurde und wenn ja weshalb (3). Wurde der durch die Heuristik adressierte Aspekt lediglich übersehen (4), wird geklärt, ob ein entsprechendes Feature vonnöten ist und somit ein Verbesserungspotenzial erkannt wurde (5) oder ob die Heuristik in diesem System ignoriert werden kann (6).

Diese Vorgehensweise erzeugt Nutzen auf zwei Seiten: während die Expert/-innen, die das Interview führen, ein ganzheitliches Systemverständnis aufbauen, werden den befragten Gestalter/-innen des Systems potenzielle Defizite und somit Verbesserungsmaßnahmen aufgezeigt.

4.2 Fallbeispiele für den Einsatz der Heuristik-basierten sozio-technischen Systemanalyse

Um die Tauglichkeit der Heuristiken zur Systemevaluation zu testen, wurden unter anderem Industrie-4.0-Szenarien am Future Work Lab⁸ des Fraunhofer-Instituts für Arbeitswirtschaft und Organisation IAO analysiert. Im Rahmen einer circa dreistündigen, durch systemkundige Mitarbeiter/-innen begleiteten, Begehung verschiedener Demonstratoren und Szenarien lernten die Autoren sechs Industrie-4.0-Lösungen kennen, die autonome Assistenz beinhalten. Die Heuristiken wurden während der Begehung genutzt, um gezielt Nachfragen zu stellen, um ein anschließendes Interview mit einem systemkundigen Mitarbeiter zu

8 <https://futureworklab.de/> vom 21.8.2018.

strukturieren und um im Nachgang Fallbeschreibungen, die sich auf die Heuristiken beziehen, zu formulieren.

Die folgenden Absätze beinhalten zwei dieser, teilweise verzahnten, Fallbeschreibungen. In Klammern wird jeweils ein Bezug zwischen den beschriebenen Erkenntnissen über das System und den Heuristiken hergestellt; an erster Stelle steht dabei das Stichwort, das die Heuristik kennzeichnet, an zweiter Stelle wird angegeben, ob das System den Aspekt ausreichend behandelt (+), ein diesbezügliches Defizit aufweist (-) oder keine genaue Aussage möglich ist (?).

Ein Maschinenbediener oder eine Instandhalterin ist im Szenario *Mobile Mehrmaschinenbedienung* für mehrere Maschinen zuständig, die Bearbeitungsschritte autonom ausführen, auch wenn der Maschinenführer nicht unmittelbar die Maschine überwacht. Über eine SmartWatch oder ein anderes mobiles Device wird er informiert, ob Ereignisse (reguläre, wie Beendigung eines Vorgangs oder Störungen) eingetreten sind, die seine Reaktion erfordern. Das kann ihn von monotoner Überwachungsarbeit entlasten (Balance, +) und die Arbeit für ihn interessanter machen. Der Einsatz eines einzelnen Maschinenbedieners wird dadurch effizienter (Effiziente Organisation, +). Potenziell kann es auch zu temporärer Überlastung (Balance, -) und Stress kommen, wenn mehrere Ereignisse, die eine Reaktion erfordern, gemeinsam auftreten. Die Meldungen können auch an mehrere Personen verteilt werden. Das Kommunikationskonzept dafür ist noch unklar (Kommunikationsunterstützung, ?), also z.B., ob delegiert werden kann, dass jemand – und wer – sich um eine Störungsanzeige kümmern soll. Auch die Rollenaufteilung zwischen den Personen ist unklar (Kommunikationsunterstützung, ?). Es ist nicht klar, ob im Fall einer Überlastung über die Smart Watch-Unterstützung angefordert werden kann (Kommunikationsunterstützung, ?; Unterstützende Technik, ?). Unklar ist auch, ob die Meldungen Hinweise zur Dringlichkeit geben (Nachvollziehbarkeit, ?; Informationsaustausch, ?). Es ist nicht klar, inwieweit die Daten für das Management zu Überprüfung der Leistung und der Rechtzeitigkeit der Reaktionen (Informationsaustausch, ?) herangezogen werden können. Über Lernmöglichkeiten ist nichts bekannt (Kom-

patibilität, -). Ebenso wenig, ob es Feedback gibt oder ob man sich mit anderen Maschinenbediener/-innen vergleichen kann (Nachvollziehbarkeit, -), indem man sieht, auf wie viele Ereignisse andere im Durchschnitt reagieren müssen. Man kann über Regeln (einfache Wenn-dann-Bestimmungen) festlegen, wer welche Informationen bekommt (Flexibilität, +). Statt einer SmartWatch kann auch ein Smartphone genutzt werden (Flexibilität, +; Unterstützende Technik, +). Insgesamt ist es zur Verteilung potenzieller Belastungsspitzen sinnvoller, mehrere Maschinenbediener/-innen für ein – dann durchaus größeres – Anlagengeld einzusetzen (Effiziente Organisation, ?). Wenn mehrere Maschinenbediener/-innen adressierbar sind, könnte die Koordination darüber hinaus erleichtert werden, indem nur die Bediener/-innen eine Nachricht erhalten, die sich zum einen räumlich am nächsten an der betroffenen Maschine aufhalten und zum anderen nicht gerade mit der Bearbeitung eines Auftrags befasst sind. Wenn jemand direkt mit der betroffenen Maschine zu Gange ist, muss in diesem Fall auch keine Nachricht verschickt werden (Kommunikationsunterstützung). Insgesamt ist dafür die Lokalisierbarkeit der Maschinenbediener/-innen zu unterstützen (Unterstützende Technik).

Die mobile Mehrmaschinenbedienung lässt sich mit dem Szenario *Nachvollziehbarkeit der Maschinenablaufhistorie* verzahnen. Angenommen der gerufene Maschinenarbeiter wendet sich einer Maschine zu, bei der eine Störung vorliegt. Er hat drei Informationsquellen, um die Störung zu analysieren: Ein Dashboard, das in erster Linie Maschinendaten im historischen Verlauf darstellt, Videos, die die Ereignisse an der Maschine, insbesondere das Werkstück zeigen, sowie eine virtuelle Repräsentation der Maschine, bei der die Teile, die mit der Störung in Verbindung stehen, besonders gekennzeichnet werden. Der Zustand der Maschine und der Weg zu diesem Zustand sind dadurch gut nachvollziehbar (Nachvollziehbarkeit, +). Dies ist insgesamt wichtig, um das Verhalten Autonomer Systeme rekonstruieren zu können, um es gegebenenfalls im Nachhinein zu verstehen.

Vorgänge können besser verstanden bzw. erlernt werden (Nachvollziehbarkeit, +; Kompatibilität, +) und Informationen zur Störungs-







behebung werden bereitgestellt (Informationsaustausch, +). Zum Beispiel können auch Explosionsdarstellungen bei der virtuellen Repräsentation angeboten werden. Der Maschinenbediener/die Maschinenbedienerin erhält Hinweise (Informationsaustausch, +), um zu entscheiden, was zu tun ist – also ob beispielsweise ein Ersatzteil einzuwechseln ist. Das Hinzuziehen einer weiteren Person an einem anderen Ort zu Zwecken der Erörterung der Vorgänge an der Maschine ist nicht vorgesehen (Kommunikationsunterstützung, -). Es ist unklar, wie man durch die Darstellungen navigieren kann, ob es also bestimmte Stellen gibt, die angesprochen werden können, um die verfügbaren Daten möglichst effizient zu sichten (Unterstützende Technik, ?; Nachvollziehbarkeit, ?). Es wäre von Interesse zu sehen, ob ein erfahrener Maschinenarbeiter/eine erfahrene Maschinenarbeiterin die Darstellungen benötigen würde (Balance, ?) bzw. nutzen will. Es kann von einer hohen Überdeckung zwischen tatsächlichen Maschinenabläufen und der technisch erzeugten Datenrepräsentation ausgegangen werden (Unterstützende Technik, +). Bei dem Szenario bleibt unklar, wie vermieden wird, dass die aufgezeichneten Daten auch zu Zwecken der Leistungs- oder Verhaltenskontrolle herangezogen werden können (Informationsaustausch, -). Es bleibt auch offen, inwieweit sich das Verhalten des Maschinenbedieners/der Maschinenbedienerin selbst in den aufgezeichneten Datenströmen widerspiegelt (Nachvollziehbarkeit, ?; Informationsaustausch, ?).

5. RELEVANZ DER HEURISTIK-GELEITETEN EXPLORATION FÜR DIE SOZIO-TECHNISCHE INTEGRATION AUTONOMER SYSTEME

Der oben beschriebene Fall (Kapitel 4.2) kann genutzt werden, um die Eigenschaften Autonomer Systeme und die Rollen, die Menschen in diesem Kontext künftig spielen, zu erörtern. Auch für eine solche Erörterung helfen die Heuristiken als Reflexionsunterstützung.

Zunächst wird eine wichtige Charakteristik deutlich: Maschinen, wie etwa Drehautomaten, entsprechen umso eher der Vorstellung eines Autonomen Systems, je mehr eine feingranulare Interaktion mit ihnen überflüssig wird. Autonome Systeme erbringen ihre Leistung und ihren Nutzen weitgehend ohne die Mitwirkung eines Bedieners und einer Nutzerin (vgl. Schmidt/Herrmann 2017). Nur in Ausnahmen (Störung) oder an bestimmten Punkten des Nutzungsverlaufs (Einrichten, Qualitätskontrolle) sind Interaktionen erforderlich, die man zum Teil als Intervention verstehen kann. Das Zusammenspiel zwischen Menschen und Autonomen Systemen lässt sich weniger als kontinuierliche Interaktion beschreiben, sondern eher unter dem Paradigma der Intervention. Solche Interventionen und dafür geeignete Nutzungsschnittstellen sind durch die in Abbildung 10 dargestellten Gesichtspunkte gekennzeichnet.

Abbildung 10: Kennzeichnende Aspekte von Interventionen und entsprechenden Nutzerschnittstelle

	1) Es gibt keinen vorher festgelegten Plan, wann und ob Intervention auftritt; sie geschieht ausnahmsweise.
	2) Eingriffe sind nur für einen begrenzten Zeitraum und für einen begrenzten Wirkungsbereich möglich.
	3) Interventionen unterstützen die explorative Erkundung der Auswirkungen von Variationen in der Vorgehensweise oder in der Konfiguration eines autonomen Systems.
	4) Menschen müssen in der Lage sein, Interventionen so schnell zu starten, indem sie technische Mittel oder Kommunikation anwenden, dass die gewünschten Wirkungen rechtzeitig eintreten.
	5) Situationen, die ein Eingreifen erfordern, sind emergent und kontingent und tragen zur Entstehung neuer Verhaltensmuster bei.
	6) Im Wechsel mit der Rekonfiguration hilft die Intervention zyklisch, das automatisierte oder routinierte Verhalten zu verbessern.

Quelle: eigene Darstellung in Anlehnung an Schmidt/Herrmann 2017

Diese Eigenschaften lassen sich anhand des in Kapitel 4.2 beschriebenen Fallbeispiels verdeutlichen:

Ad 1) Während das Einrichten der Maschine oder die Entnahme eines gefertigten Teils vorhersehbar sind, gilt dies für Störungen oder für bestimmte Wartungsaufgaben, wie etwa dem Nachschleifen eines Werkzeuges, nicht. Hier entspricht es dann einer Erhöhung der Autonomie eines Systems, wenn es Störungen und absehbare Wartungsbedarfe selbst erkennt und melden kann. Dabei muss man davon ausgehen, dass es hier zu False-positive-Meldungen kommen kann. Daher muss ein Mensch im Nachgang an eine Meldung entscheiden, wie sie zu bewerten ist und welche Maßnahmen durchzuführen sind. Dazu gehört auch, dass zunächst die Historie, die zu einer Störung oder zur Meldung eines Wartungsbedarfes geführt hat, nachvollziehbar ist. Ohne Nachvollziehbarkeit (Heuristik 1) kann die Notwendigkeit oder die Art einer benötigten Intervention nicht eingeschätzt werden. Entsprechende Möglichkeiten werden in dem Fallbeispiel durch die Nachvollziehbarkeit der Maschinenhistorie angeboten. Insgesamt werden Menschen durch solche automatisch veranlassten Meldungen – sofern sie in der Mehrheit zutreffend sind (Heuristik 8, unterstützende Technik) – entlastet, da sie nicht aufgrund mangelnder Planbarkeit der eintretenden Ereignisse in eine permanente Monitoring-Aufgabe gezwungen werden (Heuristik 5, Balance). Da Interventionen jedoch seltener sind und weniger Aufmerksamkeit erfordern als kontinuierliche Interaktion oder kontinuierliches Monitoring, können einem Maschinenbediener/einer Maschinenbedienerin mehrere Maschinen zur Betreuung zugewiesen werden, was unter Umständen bei gehäuften Meldungen oder Störungen den Stress erhöhen kann (Heuristik 2, Balance).

Ad 2) Es entspricht dem Wesen der Intervention, dass sie zeitlich begrenzt ist. Sie findet nur so lange statt, wie etwa eine Reparatur dauert. Wenn z.B. ausprobiert wird, ob eine Erhöhung des Kühlmittelzuflusses bei der gemeldeten Überhitzung eines Werkstücks Abhilfe bietet, dann ist diese Maßnahme zeitlich zu begrenzen, etwa für die Dauer eines Erprobungszeitraums oder bis zum nächsten Werkzeugwechsel.

Auch das Ausmaß der Erhöhung des Kühlmittelzuflusses wird innerhalb bestimmter Grenzen festgelegt sein.

Ad 3) Intervention beinhaltet die Möglichkeit der Erkundung aus der Perspektive einer Was-wäre-wenn-Frage. In dem obigen Beispiel wird also z.B. der Frage nachgegangen, ob es sich durch die Erhöhung des Kühlmittelzuflusses vermeiden lässt, ein Werkzeug sofort zu wechseln, was in Fällen sinnvoll ist, in denen für die Fertigstellung eines Teils, etwa in einem Drehautomaten, nur noch wenig Zeit benötigt wird. Dies entspricht jedoch immer einer experimentellen Vorgehensweise, über deren Angemessenheit ein Mensch entscheidet, der hierfür durch entsprechende organisatorische Regelungen über die entsprechende Autonomie bzw. Entscheidungsfreiheit verfügen muss (Heuristik 2, Flexibilität; Heuristik 6, Kompatibilität). Hier wird deutlich, dass die Erhöhung der Autonomie eines technischen Systems nicht notwendigerweise die Reduzierung der Autonomie der beteiligten menschlichen Akteur/-innen impliziert. Diese hängt vielmehr von den Interventionsmöglichkeiten und den organisatorischen Regelungen ab.

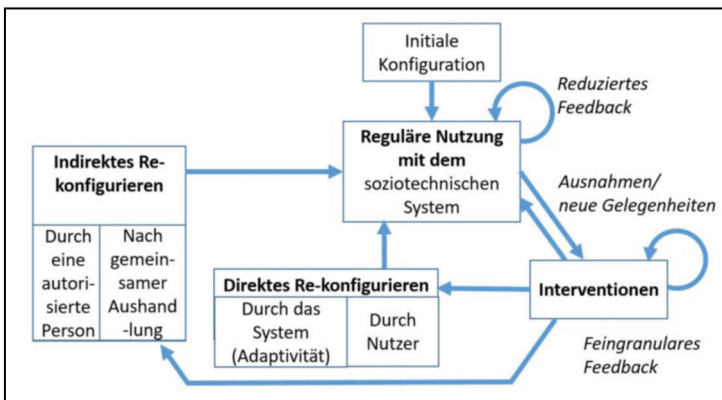
Ad 4) In dem obigen Fallbeispiel wird unterstellt, dass eine Benachrichtigung eines räumlich nicht unmittelbar an der Maschine tätigen Menschen ausreichend ist, auch wenn bis zum Eingriff dann noch Zeit vergeht. Dies trifft nicht auf alle Konstellationen zu, da gegebenenfalls eine sofortige Maßnahme erfolgen sollte, etwa durch eine Fernbedienung oder indem man jemanden informiert, der sich in der unmittelbaren Nähe befindet. Hier wird deutlich, dass die technische Ausgestaltung (Heuristik 8, unterstützende Technik) unmittelbare Eingriffsmöglichkeiten eröffnen muss oder dass – aus einem übergeordneten sozio-technischen Blickwinkel – die Zusammenarbeit mit anderen Mitarbeiter/-innen so koordiniert ist (Heuristik 3, Kommunikationsunterstützung), dass diese eine Intervention übernehmen können.

Ad 5) Das Angebot von Intervention berücksichtigt, dass es im Umgang mit technischen Systemen immer wieder zu emergenten bzw. kontingenten (vgl. Pedersen 2000) Zustandsveränderungen kommen kann, auf die reagiert werden muss. Autonome Systeme eröffnen selbst durch ihre Interaktion mit der Umwelt eine Erhöhung von Komplexität

– im obigen Fallbeispiel etwa dadurch, dass ein Maschinenbediener/ eine Maschinenbedienerin mehr Maschinen betreuen kann, als dies vorher der Fall war. Das kann potenziell dazu führen, dass in nicht planbarer Weise so viele Meldungen eingehen, dass eine Prioritäten- setzung erforderlich ist. Der Umgang mit solchen emergenten Situa- tionen fordert bzw. eröffnet Möglichkeiten der Kompetenzentwicklung auf Seiten der Mitarbeiter/-innen, die wiederum die Kompatibilität zwischen sozio-technischem System und der zu bewältigenden Anfor- derungen erhöht (Heuristik 6, Kompatibilität).

Ad 6) Wenn im Zuge der ermöglichten Lernprozesse deutlich wird, dass bestimmte Arten der Intervention wiederholt notwendig werden, dann gibt dies Anlass zur Weiterentwicklung (Heuristik 2, Flexibilität) des Systems, welche im Bereich der technischen und/oder organisatori- schen Regelungen liegen kann. Abbildung 11 verdeutlicht, wie Inter- vention und Konfiguration zusammenspielen können.

Abbildung 11: Zyklus von Intervention und Re-Konfiguration



Quelle: eigene Darstellung in Anlehnung an Schmidt/Herrmann 2017: 42

Hier wird auch deutlich, dass es sinnvoll ist, einen engen Zusammen- hang zwischen Autonomie und evolutionärer Weiterentwicklung zu unterstellen. Allerdings ist es zur Unterstützung der Autonomie nicht

nur notwendig, den Hintergrund einer ad hoc eingetretenen Situation zu durchschauen, wie es in dem obigen Fallbeispiel vorgesehen ist, sondern auch Informationen so bereitzustellen (Heuristik 4, Informationsaustausch), dass sie einen längeren Zeitraum abdecken, um anhand einer Vielzahl von Fällen zu beurteilen, ob eine Re-Konfiguration sinnvoll ist.

Im Hinblick auf die Rollenverteilung zwischen Mensch und Autonomem System wird deutlich, dass es von ausschlaggebender Bedeutung ist, ob die Anpassung des Autonomen Systems von Menschen vollzogen werden kann bzw. muss oder ob die Anpassung letztlich durch eine eigene Adaption der Systeme, etwa mithilfe von maschinellem Lernen stattfindet. Die Beantwortung der Frage, ob hier Anpassung durch maschinelles Lernen möglich ist, hängt davon ab, ob sich große Mengen von Daten zu ausreichend vergleichbaren Situationen gewinnen lassen. Maschinelles Lernen ist skalierbar; Daten, die an verschiedenen Instanzen desselben Maschinentyps gesammelt werden, lassen sich aggregieren und das aus diesen Daten erzeugte Wissen ist prinzipiell für alle Instanzen nutzbar. Je ausgeprägter aber eine Maschine als Autonomes System ihre eigenen Besonderheiten hat (etwa durch ihre Wartungsgeschichte), desto schwerer ist es, ausreichend skalierbare Trainingsdaten für das maschinelle Lernen zu generieren – und desto eher müssen menschliche Akteur/-innen im Verlauf von Intervention, Erprobung und Re-Konfiguration aktiv werden.

Am obigen Fallbeispiel wird ebenfalls deutlich, dass mit der Erhöhung der Autonomie auch geänderte Kooperations- und Koordinationsmöglichkeiten einhergehen. Meldungen müssen nicht notwendigerweise nur an einen Bediener/eine Bedienerin gerichtet sein, sondern können an ein Team adressiert werden. Hier ergeben sich dann Entscheidungsanforderungen, wer welche Aufgaben übernimmt und wo dabei welches Kompetenzspektrum am besten zum Einsatz kommt. Prinzipiell können solche Verteilungsaufgaben wiederum durch den Einsatz von Systemen unterstützt werden, die auf Künstlicher Intelligenz basieren. Hier geht es dann auch um die Berücksichtigung interindividueller Unterschiede und im Zeitverlauf auch intraindividueller

Differenzen und Präferenzen (Heuristik 5, Balance). Die Herausforderung, hier geeignete Daten in großer Zahl zu gewinnen, die das Training maschinellen Lernens unterstützen, ist dann noch höher. Folglich erscheint es naheliegend, dass koordinative Entscheidungsaufgaben zunächst im menschlichen Bereich verbleiben.

6. FAZIT UND KOMMENDE ARBEITSSCHRITTE

Das Zuordnungsexperiment zeigt, dass das vorgestellte Heuristik-Set zumindest die gesammelten sozio-technischen Problemfälle nahezu komplett abdeckt. Es ist jedoch zu beachten, dass diese Problemsammlung schon vorgegeben war und aufgrund intensiver empirischer Evaluation und/oder diskursiver Erörterungen entstand. Es ist noch ungeklärt, ob mithilfe der diskutierten Heuristiken die gleichen Probleme effektiv und effizient entdeckt worden wären.

Ein wesentlicher Erfolgsfaktor für die Problemidentifikation mittels Heuristiken ist deren Formulierung bzw. intuitive Verständlichkeit. Bisher wurden die Heuristiken vorrangig durch die an deren Entwicklung beteiligten Expert/-innen eingesetzt, welche im Laufe der Arbeit ein gutes Verständnis der durch die Heuristiken referenzierten Aspekte der Systemgestaltung entwickelt haben. Das bisherige – auf 50 Sub-Heuristiken aufbauende – Design erscheint wenig handhabbar für Dritte.

Wir verstehen die Herausforderung einer Heuristik-basierten sozio-technischen Evaluationsmethode nicht als endgültig lösbar. Vielmehr erwarten wir, dass die Heuristiken in der Praxis noch weiterzuentwickeln sind. Input für ein erstes Redesign waren die gewonnen Erkenntnisse aus dem Einsatz der Heuristiken (z.B. deren Verständlichkeit und Handhabbarkeit) und dem Zuordnungsexperiment.

Unabhängig von diesen Limitierungen zeigten erste Tests der Heuristiken bei Praxispartner/-innen des Projekts *Heuristiken für die Industrie 4.0*, dass sie vielfältig einsetzbar sind. Sie helfen, einen Großteil der kritischen Aspekte sozio-technischer Systemgestaltung zu ad-

ressieren. Der betriebliche Alltag der Praxispartner/-innen verhinderte teilweise, dass das zu untersuchende System von Expert/-innen beobachtet werden konnte und Teilnehmer/-innen des Systems befragt werden konnten. Aber auch bei dem zunächst als Notlösung konzipierten Ansatz, die Analyse auf Basis eines Interviews mit Designer/-innen des Systems durchzuführen, halfen die Heuristiken, eine fundierte Diskussion anzuregen (vgl. Kapitel 4.1). Während die Interviewer ein, für den begrenzten Zeitrahmen des Interviews, weitreichendes Systemverständnis aufbauen konnten, profitierten die interviewten Designer/-innen von der Aufdeckung von Verbesserungspotenzialen. Die Heuristiken können auch für die Strukturierung der Dokumentation von sozio-technischen Lösungen eingesetzt werden. Außerdem müssen sie nicht nur auf bereits bestehende Systeme angewandt werden, sondern können auch den Designprozess von Anfang an begleiten.

Im Hinblick auf die Aufgabengestaltung für Menschen, die mit Autonomen Systemen umgehen, wird deutlich, dass über die Wahrnehmungen von Interventionen, insbesondere bei der Nutzung mehrerer Systeme, reichhaltige und komplexe Aufgaben zu erwarten sind. Nachvollziehbarkeit (Heuristik 1) sowie flexible Erkundung und Exploration (Heuristik 2) sind jedoch wichtige Voraussetzungen, damit die Notwendigkeit oder die Art einer benötigten Intervention eingeschätzt werden können.

Insgesamt werden Menschen durch Autonome Systeme – etwa durch automatische Meldungen – entlastet, da sie nicht aufgrund mangelnder Planbarkeit der eintretenden Ereignisse in eine permanente Monitoring-Aufgabe gezwungen werden. Allerdings bestehen hohe Anforderungen, das Verhalten des Systems mit den Anforderungen des Umfeldes und dessen Dynamik kompatibel zu halten. Sofern diese Kompatibilität nicht gewährleistet ist, wird eine Fehlbeanspruchung durch Stress wahrscheinlicher (Heuristik 5, Balance). Autonome Systeme erfordern auch eine Autonomie auf der menschlichen Seite, wenn man davon ausgeht, dass die Anpassungsprozesse zu Erzielung von Kompatibilität durch experimentelles Explorieren gestützt werden, das von menschlichen Akteur/-innen autonom verantwortet werden muss.

Die jeweils notwendige Entscheidungsfreiheit ist in der Regel in die Koordination der Zusammenarbeit innerhalb von Gruppen einzubinden. Die Übernahme von Intervention kann dabei flexibel verteilt werden. Dabei entstehen wiederum neue Aufgabentypen der Koordination.

Intervention und Konfiguration bedingen sich gegenseitig (siehe Abbildung 11). Es ergibt sich ein enger Zusammenhang zwischen Autonomie und evolutionärer Weiterentwicklung, für den die beteiligten Mitarbeiter/-innen ausreichend vorzubereiten und in die entsprechenden Informationsflüsse einzubinden sind. Für die Zukunft der Aufgabenverteilung ist es ausschlaggebend, inwieweit die Anpassung des Autonomen Systems von Menschen vollzogen werden kann bzw. muss oder ob die Anpassung letztlich durch eine eigene Adaption der Systeme, etwa mithilfe von maschinellem Lernen stattfindet.

LITERATUR

- Ackermann, Mark/Prilla, Michael/Stary, Christian/Herrmann, Thomas (Hg.), *Designing Healthcare that Works. A Socio-Technical Approach*, London/San Diego u.a.: Academic Press.
- Baker, Kevin/Greenberg, Saul/Gutwin, Carl (2001): „Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration“, in: Murray Reed Little/Laurence Nigay (Hg), *Engineering for Human-Computer Interaction*, Berlin/Heidelberg: Springer, S. 123-139.
- Baxter, Gordon/Sommerville, Ian (2011): „Socio-Technical Systems. From Design Methods to Systems Engineering“, in: *Interacting with Computers* 23/1, S. 4-17.
- Brockhaus Enzyklopädie (1995): *Brockhaus Enzyklopädie. Ergänzungsbände, Deutsches Wörterbuch in drei Bänden*, Bd. 27: *Deutsches Wörterbuch Gluc-Reg*, Mannheim: Wissenmedia, S. 1568.
- Cherns, Albert (1987): „Principles of Sociotechnical Design Revisited“, in: *Human Relations* 40/3, S. 153-161.
- Clegg, Chris W. (2000): „Sociotechnical Principles for System Design“, in: *Applied Ergonomics* 31/5, S. 463-477.

- Clement, Andrew (1993): „Considering Privacy in the Development of Multi-Media Communications“, in: Computer Supported Cooperative Work 2/1-2, S. 67-88.
- Dix, Alan/Finlay, Janet/Abowd, Gregory/Beale, Russel (2004): Human-Computer Interaction, Harlow/New York: Pearson Education.
- Dunckel, Heiner (1989): „Arbeitspsychologische Kriterien zur Beurteilung und Gestaltung von Arbeitsaufgaben im Zusammenhang mit EDV-Systemen“, in: Susanne Maaß/Horst Oberquelle (Hg.), Software-Ergonomie'89. Aufgabenorientierte Systemgestaltung und Funktionalität, Stuttgart: B.G. Teubner, S. 69-79.
- Eason, Ken (1988): Information Technology and Organisational Change, Boca Raton: CRC Press.
- Fischer, Gerhard/Herrmann, Thomas (2011): „Socio-Technical Systems. A Meta-Design Perspective“, in: International Journal of Sociotechnology and Knowledge Development 3/1, S. 1-33.
- Greenberg, Saul/Fitzpatrick, Geraldine/Gutwin, Carl/Kaplan, Simon (1999): „Adapting the Locales Framework for Heuristic Evaluation of Groupware“, in: Proceedings of OZCHI'99 Australian Conference on Computer Human Interaction, Wagga Wagga, 28.-30.11. 1999, Brisbane: Charles Sturt University, S. 30-36.
- Grote, Gudela (2015): „Gestaltungsansätze für das komplementäre Zusammenwirken von Mensch und Technik in Industrie 4.0“, in: Hartmut Hirsch-Kreinsen/Peter Ittermann/Jonathan Niehaus (Hg.), Digitalisierung industrieller Arbeit: Die Vision Industrie 4.0 und ihre sozialen Herausforderungen, Baden-Baden: Nomos, S. 132-147.
- Grote, Gudela/Ryser, Cornelia/Wäfler, Toni/Windischer, Anna/Weik, Steffen (2000): „KOMPASS. A Method for Complementary Function Allocation in Automated Work Systems“, in: International Journal of Human-Computer Studies 52/2, S. 267-287.
- Hackman, J. Richard/Oldham, Greg, R. (1975): „Development of the Job Diagnostic Survey“, in: Journal of Applied Psychology 60/2, S. 159-170.

- Herrmann, Thomas (2012): Kreatives Prozessdesign. Konzepte und Methoden zur Integration von Prozessorganisation, Technik und Arbeitsgestaltung, Heidelberg: Springer.
- Herrmann, Thomas/Ackermann, Mark S./Goggins, Sean P./Stary, Christian/Prilla, Michael (2017): „Designing Health Care that Works – Socio-Technical Conclusions“, in: Mark Ackermann/Michael Prilla/Christian Stary/Thomas Herrmann (Hg.), Designing Healthcare that Works. A Socio-Technical Approach, London/San Diego u.a.: Academic Press, S. 187-203.
- Herrmann, Thomas/Prilla, Michael/Nolte, Alexander (2016): „Socio-Technical Process Design – The Case of Coordinated Service Delivery for Elderly People“, in: Fabrizio D’Ascenzo/Massimo Magni/Alessandra Lazazzara/Stefano Za (Hg.), Blurring the Boundaries Through Digital Innovation, Cham: Springer, S. 217-229.
- Herrmann, Thomas/Wulf, Volker/Hartmann, Anja (1996): „Requirements for the Human Centred Design of Groupware“, in: The Design of Computer Supported Cooperative Work and Groupware Systems 12, S. 77-99.
- International Organisation for Standardization (ISO) (2006): ISO 9241-110:2006. Ergonomics of Human-System Interaction, Teil 110: Dialogue principles, <https://www.iso.org/standard/38009.html> vom 1.4.2018.
- Mumford, Enid (1983): Designing Human Systems for New Technology. The ETHICS Method. Manchester: Manchester Business School.
- Nielsen, Jakob (1994): „Enhancing the Explanatory Power of Usability Heuristics“, in: Beth Adelson/Susan Dumais/Judith Olson (Hg.), Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York: ACM, S. 152-158.
- Nielsen, Jakob/Molich, Rolf (1990): „Heuristic Evaluation of User Interfaces“, in: Beth Adelson/Susan Dumais/Judith Olson (Hg.), Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York: ACM, S. 249-256.

- Nierhoff, Jan/Herrmann, Thomas (2017): „Data Elicitation for Continuous Awareness of Team Climate Characteristics to Improve Organizations’ Creativity“, in: Proceedings of the 50th Hawaii International Conference on System Sciences, HICSS 2017, Waikoloa, Hawaii: University of Hawaii, S. 204-213.
- Pedersen, Poul Poder (2000): „Our Present: Postmodern?“, in Heine Andersen/Lars B. Kaspersen (Hg.), Classical and Modern Social Theory, Oxford: Blackwell Publishers. S. 412-431.
- Prilla, Michael/Herrmann, Thomas (2017): „Challenges for Socio-Technical Design in Health Care: Lessons Learned from Designing Reflection Support“, in: Mark Ackermann/Michael Prilla/Christian Stary/Thomas Herrmann (Hg.), Designing Healthcare that Works. A Socio-Technical Approach, London/San Diego u.a.: Academic Press, S. 149–166.
- Reijers, Hajo A./Mansar, Selma L. (2005): „Best Practices in Business Process Redesign. An Overview and Qualitative Evaluation of Successful Redesign Heuristics“, in: Omega 33/4, S. 283-306.
- Rost, Martin/Bock, Kirsten (2011): „Privacy by Design and the New Protection Goals“, in: Datenschutz und Datensicherheit 1, S. 30-35.
- Schafner, Marlene/Lacueva, Francisco, J./Hannola, Lea/Damalas, Stelios A./Nierhoff, Jan/Herrmann, Thomas (2018): „Insights into the Introduction of Digital Interventions at the Shop Floor“, in: Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference, New York, ACM, S. 331-338
- Schmidt, Albrecht/Herrmann, Thomas (2017): „Intervention User Interfaces: A New Interaction Paradigm for Automated Systems“, in: Interactions 24.5, S. 40-45.
- Schneider, Wolfgang (2008): Ergonomische Gestaltung von Benutzungsschnittstellen. Kommentar zur Grundsatznorm DIN EN ISO 9241-110, Berlin: Beuth.
- Shneiderman, Ben/Plaisant, Catherine/Cohen, Maxine/Jacobs, Steven (2010): Designing the User interface. Strategies for Effective Human-Computer Interaction, London: Pearson Education.

- Stangl, Werner (2018): Stichwort: ‚Heuristik‘, in: Online Lexikon für Psychologie und Pädagogik, <http://lexikon.stangl.eu/1963/heuristik/> vom 24.7.2018.
- Suchman, Lucy (1995): „Making Work Visible“, in: Communications of the ACM 38/9, S. 56-64.
- Wessells, Michael G. (1982): Cognitive Psychology, New York: Harper & Row.
- Zimbardo, Philip G. (2004). Psychologie, Hallbergmoos: Pearson Studium.

DANKSAGUNG

Wir danken Prof. Dr. Isa Jahnke, Prof. Dr. Alexander Nolte und Dr. Rainer Skrotzki für die Mitarbeit bei der Gruppierung der Heuristiken und für die inhaltliche Erörterung und die wertvollen Hinweise, die die Ausarbeitung des hier untersuchten Sets von Heuristiken entscheidend unterstützten.

Historische und gesellschaftliche Perspektiven

Technik und Autonomie

Kulturhistorische Bemerkungen zu einem komplexen Verhältnis

Martina Heßler

Im Jahr 1960 baute der Schweizer Künstler Jean Tinguely eine riesige Maschine, die unter anderem aus Rädern, Motoren, Kinderwagen und einem Klavier bestand. Es handelte sich um einen Zeichenautomat. Tinguely hatte bereits zuvor Zeichenautomaten konstruiert, die beispielsweise den Stil von Jackson Pollock imitierten. Die Maschine von 1960 unterschied sich allerdings von den früheren Automaten, denn sie hatte einen besonderen Zweck: Sie sollte sich selbst zerstören.¹

Während einer spektakulären 27-minütigen Aktion im New Yorker Museum of Modern Art verweigerte Tinguely jeden menschlichen Eingriff, der die Zerstörung hätte aufhalten können. Er entschied sich also für die Devise *Humans Out of the Loop*, wie es heute im Kontext autonomer Maschinen formuliert wird. Tinguelys Maschine war allerdings noch keine autonome Technik im heutigen Sinne, sondern lediglich ein Automat, der selbsttätig und ohne menschlichen Eingriff funktionierte, ein Automat, der einmal in Gang gesetzt, sein zuvor gesetztes Pro-

1 Vgl. zur Kunstaktion: <https://www.tinguely.ch/de/ausstellungen/ausstellungen/2010/Rotozaza.html>.

gramm ablaufen ließ. Während der Performance geschah jedoch Unvorhergesehenes: Das Klavier fing an zu brennen, ein Feuerwehrmann musste unerwartet eingreifen, um die Situation nicht außer Kontrolle geraten zu lassen.

Tinguelys Installation wirft Fragen auf, die heute weitaus dringlicher erscheinen als 1960. Er thematisierte insbesondere die Frage nach der Position der Menschen in automatisierten Prozessen. Tinguely verweigerte den Eingriff der Menschen. Die Maschine sollte gleichsam von selbst funktionieren und ihr Ziel erreichen. Der im automatisierten Ablauf nicht vorgesehene Brand des Klaviers, ein Störfall, machte den menschlichen Eingriff schließlich aber unabdingbar. Dies verweist wiederum auf die Forderungen, der Mensch müsse letztlich immer einen Notknopf drücken oder ‚den Stecker ziehen‘ können. Diese Forderungen werden insbesondere im Kontext autonomer Technik gestellt. Welche Rolle ist den Menschen nun gegenüber autonomer Technik zuzuweisen? Hätte eine autonome, intelligente Technik flexibel und situationsangepasst reagiert? Hätte sie mit dem unvorhergesehenen Ereignis des Brandes umgehen können?

Tinguelys Installation evoziert darüber hinaus weitere, im heutigen Kontext zentrale Fragen. So ist die Maschine ein Absurdum: ein Zeichenautomat, der sich selbst zerstört, ohne innezuhalten – ein sinnloses Unterfangen. Die Maschine kann dabei weder ihr eigenes Tun noch ihren Zweck beurteilen und hinterfragen. Dies gilt gleichermaßen für heutige autonome Technik. In der philosophischen Tradition ist der Autonomiebegriff jedoch genau mit dieser menschlichen Fähigkeit der Urteilstkraft verknüpft. In dieser Tradition des Autonomiebegriffs steht insbesondere die kulturkritische Debatte und Technikkritik. Heute spielt sie weniger eine Rolle, indem der Technik nun Autonomie zugesprochen wird, allerdings in verschobenem Sinne, wie im Folgenden gezeigt werden soll.

Begriffliches: Autonomie der Technik?

Dem gegenwärtigen Begriff der *autonomen Technik* ist eine neue Bedeutung von Autonomie inhärent, wie er zugleich eine neue Dimension im Mensch-Maschinen-Verhältnis anzeigt. Die Rede ist nicht von automatisierter Technik oder von Automaten, also von selbsttätig, ohne menschlichen Eingriff ablaufenden, jedoch von Menschen programmierten technischen Prozessen, sondern von *autonomer* Technik. Damit ist die Referenz auf den Autonomiebegriff hergestellt.

In der griechischen Antike wurde der Begriff der Autonomie im politischen Zusammenhang verwendet. Als zentrale politische Kategorie bezeichnete Autonomie bei Herodot „innere und äußere politische Freiheit (ἐλευθερία) im Gegensatz zu einer Abhängigkeit von Fremdherrschaft und im Gegensatz zur inneren Staatsform der Tyrannis“ (Pohlmann: 2010: 701).

Im Mittelalter war der Begriff der Autonomie unbekannt. Im 17. und 18. Jahrhundert findet er sich in der Jurisprudenz. Vor allem aber Immanuel Kant prägte den Begriff im Sinn der Eigen-Gesetzlichkeit und der Selbstbestimmung. Mit Kant wurde der Begriff an Menschen gebunden *und* an den Vernunftgebrauch. Autonomie meinte Selbstbestimmung durch Vernunft. Damit einher ging das Konzept des Subjekts, der Souveränität und der Willens- und Entscheidungsfreiheit, unabhängig von Äußerem (vgl. Pohlmann 2010). Diese Bindung von Autonomie an das menschliche Subjekt beeinflusste die Diskussion um Technik vom ersten Drittel des 20. Jahrhunderts bis heute, wie im Folgenden skizziert wird.

Die Debatte um autonome Technik ist immer auch eine Debatte um die Position der Menschen gegenüber Technik, zumal der Begriff *autonome Technik* stets die Unabhängigkeit *der Technik* von Menschen impliziert: Sie funktioniert ohne Menschen, das ist ihr Kennzeichen. Dies gilt gleichermaßen für den Begriff des Automaten und bedingt auch für den der Automatisierung.

Im Folgenden werden unterschiedliche historische Konzepte des Begriffs autonomer Technik im 20. und 21. Jahrhundert betrachtet. Der Beitrag schildert im ersten Teil die kulturkritischen Debatten um die *Autonomie* der Technik bis in die 1970/80er Jahre hinein, in der Technik als autonom interpretiert und als Bedrohung der Menschen wahrgenommen wurde. Im zweiten Teil wird es um Automatisierung gehen, die in anderer Weise diskutiert wurde, jedoch eine wichtige Stufe auf dem Weg zur autonomen Technik darstellt und gleichfalls eine Herausforderung für die Position der Menschen im Arbeitsprozess bedeutete. Im dritten Teil wird die Frage nach dem verwendeten Autonomiebegriff hinsichtlich heutiger autonomer Technik gestellt, um viertens mit einem Blick auf aktuelle theoretische Diskurse zu schließen, die einerseits die kulturkritischen Diskurse des 20. Jahrhunderts aufnehmen, diese jedoch in Anthropozentrismus-kritischer Manier umwenden und die Frage der autonomen Technik so neu diskutieren. Gefragt wird also erstens nach dem Wandel des Autonomiebegriffs im Verhältnis zu Technik. Zweitens stellt dabei die Frage nach der jeweiligen Position der Menschen den zweiten roten Faden des Beitrags dar.

1. AUTONOME TECHNIK: KULTURKRITISCHE PERSPEKTIVEN

Technik, so ein gängiges Narrativ der Moderne, sei ein Mittel, um die Gestaltungsmacht der Menschen herzustellen, zu erhalten und zu steigern. Technik ist demnach ein Mittel im Dienste der Menschen, ein Werkzeug, das sie selbst geschaffen haben und das sie beherrschen.

Schon früh, bereits in der Frühen Neuzeit, findet sich jedoch die Vorstellung, dass das vom Menschen Geschaffene sich verselbständigen und unabhängig von menschlichen Einflüssen werden und sich dann gar – gleichsam intentional – gegen sie wenden könne, dass es also zu einer autonomen, von Menschen unabhängigen Macht werde.

Diese Vorstellung, dass das vom Mensch Geschaffene sich verselbständigt und gegen sie kehrt, findet sich bereits in der Figur des Golem.

Der Golem ist bis heute die Metapher für die Angst, dass etwas Selbstgeschaffenes, das die Macht der Menschen vergrößern soll, eigenmächtig, unkontrollierbar und somit zu einer Bedrohung wird.

Die Geschichte des Golems reicht weit zurück bis in talmudische Zeugnisse. In der Legende des Rabbi Loew von Prag (1609) begann eine Golem-Tradition, die den Golem als willenlose und gehorsame Schöpfung der Menschen dachte (vgl. Völker 1994: 434f.). Seit dem 18. Jahrhundert wurde die gefährliche Seite des Golem thematisiert, seine übernatürliche Stärke und die Möglichkeit, dass er außer Kontrolle geraten kann (vgl. ebd.: 436f.). Der Golem stellt mithin die Urfigur der sich verselbständigenden, autonomen und nicht kontrollierbaren Technik dar, auf die im 19. und im 20. Jahrhundert häufig rekurriert wird. Seit dem frühen 19. Jahrhundert wurde die Figur des Frankenstein zu einer ähnlichen Metapher, nun explizit auf Wissenschaft und Technik bezogen.

Wie Langdon Winner in seinem wegweisenden Buch *Autonomous Technology* zusammenfasste, ist die Vorstellung, Technik sei autonom im Sinne ihrer unkontrollierbaren Unabhängigkeit von Menschen, ein typisches Konzept der Moderne, das Stress und Unbehagen auslöste. Diese Bedenken finden sich seit dem 19. Jahrhundert in der Literatur, der Kunst, der Philosophie, der Sozialwissenschaft und in der öffentlichen Meinung (vgl. Winner 1977: 13), wenngleich der Begriff der Autonomie der Technik erst in der zweiten Hälfte des 20. Jahrhunderts systematisch verwendet wurde.

In den 1950er und 1960er Jahren entwickelte sich eine Debatte um die Autonomie der Technik, in der explizit mit diesem Begriff operiert wurde. Zuvor finden sich in der philosophischen Diskussion ähnliche Gedanken, jedoch unter dem Begriff der Technik als Schicksal oder Macht. Oswald Spengler betonte in diesem Sinne die Einflusslosigkeit der Menschen: „In der Tat vermögen weder die Köpfe noch die Hände etwas an dem Schicksal der Maschinenteknik zu ändern, die sich aus innerer, seelenhafter Notwendigkeit entwickelt hat.“ (Spengler 1931: 74) Des Weiteren bediente auch er sich, wie viele Zeitgenossen, der Herr-und-Knecht- oder Herr-und-Sklaven-Metapher. Der „Herr der

Welt wird zum Sklaven der Maschine“ (ebd.: 75). Ein weiterer typischer Topos im Kontext des Autonomie-Diskurses ist der der Totalität des Technischen. Technik sei zu einer Totalität geworden, der alles untergeordnet werde (vgl. ebd.: 79).

Diese Denkweise wurde auch unter dem Konzept des Technikdeterminismus zusammengefasst, das heute als antiquiert und überwunden gilt, jedoch ungemein wirkmächtig war. Technikdeterministisches Denken bestimmte über Jahrzehnte die Vorstellungen von Technik. Noch 1985 schreiben Donald MacKenzie und Judy Wajcman, Technikdeterminismus sei die dominante Theorie, um die Beziehung zwischen Technik und Gesellschaft zu erklären (vgl. MacKenzie/Wajcman 1985: 4). Technikdeterminismus, so die beiden weiter, gehe davon aus, dass „technological change is in some sense autonomous, outside the society“ (ebd.). Die zweite wesentliche Prämisse sei, dass technischer Wandel sozialen Wandel verursache (vgl. ebd.: 5).

Das Schlagwort der Autonomie der Technik findet sich in vielen Texten der Zeit, teils weniger, teils stärker reflektiert. Insbesondere Jacques Ellul und Helmut Schelsky sind zentral für diesen Diskurs, weshalb sie hier exemplarisch näher betrachtet werden sollen.

Der Franzose Jaques Ellul publizierte 1954 sein Buch *La Technique ou l'enjeu du siècle*, das 1964 ins Englische übersetzt wurde (vgl. Ellul 1964). Ellul entwarf in diesem Buch eine Theorie der technischen Kultur, in der er Technik als entscheidende Macht moderner Entwicklungen beschrieb. Mit dem für sein Denken zentralen Begriff der Autonomie der Technik charakterisierte er Technik als eigengesetzlich. Technik folge nicht politischen, sozialen oder ökonomischen Bedingungen, sie sei auch nicht, wie die Philosophen es glauben machten, das Ergebnis menschlicher Ideen. Technik entwickle sich vielmehr völlig unabhängig und abgekoppelt von moralischen und gesellschaftlichen Werten. Technik ist in Elluls Beschreibungen eine autonome, von anderen Faktoren und von ihrer Umwelt unabhängige Sphäre, die jedoch die Umwelt maßgeblich prägt. Wie John M. Staudenmeier zusammenfasste: Technik wird für Ellul zu einer Umwelt, zu einem Milieu (vgl. Staudenmaier 1985: 137).

Elluls Denken kreist um diese Vorstellung der Autonomie der Technik. Die Autonomie der Technik sei zugleich die einzige Bedingung der Technik selbst: Technik werde durch nichts bedingt als durch sich selbst. Der Begriff der Technik meint hierbei nicht nur Maschinen oder Artefakte, sondern viel fundamentaler das *Technische*, mithin eine Logik des Vorgehens, die an Effizienz und Regelmäßigkeit orientiert ist. *La Technique* bezeichnet rationelle Methoden und Effizienz-Denken.

Ganz ähnlich sprach Schelsky, der sich auf Ellul berief, von der Sachgesetzlichkeit von Technik. Der Mensch habe keine Möglichkeit, diese „zu manipulieren oder auch nur zu überdenken“ (Schelsky 1961: 17). Der technische Sachzwang herrsche. Schelsky thematisiert vor allem den Verlust demokratischer und genuin politischer Entscheidungen. An deren Stelle sei ein an Effizienz und technischer Logik („Funktions- und Leistungsfähigkeit“, ebd.: 25) im Sinne des one best way orientierter Entscheidungszwang getreten. Der Spielraum politischer Entscheidungen nehme ab. Damit herrsche im Prinzip niemand mehr, „sondern hier läuft eine Apparatur, die sachgemäß bedient werden will“ (ebd.: 26). Er sprach von einer „Apparaturgesetzlichkeit“ (ebd.). Ganz in der Logik der Golem-Metapher formulierte er: „Der Mensch ist den Zwängen unterworfen, die er selbst als eine Welt und sein Wesen produziert.“ (Ebd.: 18)

Fasst man die zentralen Punkte dieses kulturkritischen, technikdeterministischen Diskurses um Autonomie der Technik zusammen, die hier, exemplarisch, insbesondere an Ellul und Schelsky veranschaulicht wurden, lassen sich sechs Grundannahmen benennen: *Erstens*, dass Technik die zentrale Ursache für gesellschaftlichen, kulturellen, politischen und ökonomischen Wandel darstelle. *Zweitens*, dass Technik dabei ihrer eigenen Logik, mithin einer Eigengesetzlichkeit folge; dies sei eine Logik des Funktionierens (vgl. Staudenmaier 1985: 137). *Drittens*, untrennbar damit verbunden, dass technikexterne Faktoren keinen Einfluss auf die technische Entwicklung haben. Menschen könnten die Technikentwicklung in dieser Logik nicht beeinflussen. Sie sei unabhängig und damit autonom. Menschen haben daher, *viertens*, keine an-

dere Wahl, als sich an die technische Entwicklung anzupassen. *Fünftens* sei damit das Verhältnis von Herr und Knecht, die Vorstellung, Menschen seien die Herrscher der Technik, umgekehrt. *Sechstens* schließlich bedeute dies, dass Technik sich selbst erhalte und Selbstzweck sei. Technik habe sich in dieser Logik verselbständigt. Die Menschen, als einzelne und als Gesellschaft, seien ihr ausgeliefert.

Die Prämisse dieser Argumentation ist anthropozentrisch. Die Position der Menschen wird zentral, machtvoll und gestaltend gesetzt. Aus dieser anthropozentrischen Perspektive wird die menschliche Position durch Technik hochgradig bedroht. Diese Konzepte denken Mensch und Technik dichotom, wobei der Begriff der Autonomie dem Menschen vorbehalten sein sollte. Ellul formulierte eindeutig: „There can be no human autonomy in the face of technical autonomy.“ (Ellul 1964: 138) Mensch und Technik werden gegeneinander ausgespielt; nur einer kann autonom sein. Ist die Technik autonom, ist der Mensch entmachtet bzw. Technik raubt den Menschen Entscheidungs- und Einflussmöglichkeiten.

Der Begriff der Autonomie betont mithin insbesondere die Eigengesetzlichkeit und Unbeeinflussbarkeit von Technik. Indem Autonomie aber als ausschließlich den Menschen zustehend interpretiert wird, entsteht ein Machtkampf: Entweder herrscht der Mensch *oder* die Technik, ein polares Denken, das Mensch und Technik im Modus des Herr-und-Knecht-Verhältnisses interpretiert. Autonomie wird letztlich an Macht und Beherrschung gekoppelt, an die Möglichkeit der Entscheidung und Einflussnahme. Die Kulturkritik verfehlt dabei in ihrer Technikangst jedoch den Kern des Kant'schen Autonomiebegriffs. Sie schreibt der Technik Autonomie zu, ohne zu diskutieren, dass der scheinbar autonomen Technik der Kern des Kant'schen Autonomieverständnisses fehlt: das Urteilsvermögen. Autonomie wurde hier vorrangig in den Kategorien der Macht gedacht.

Nun könnte man diese Diskurse des 20. Jahrhunderts für obsolet halten, für historisch, zumal sich seit den 1980er Jahren ein Diskurs etablierte, der in Abgrenzung hierzu stets die Gestalt- und Beeinflussbarkeit der Technik hervorhob. Gleichwohl finden sich auch heute

vielfach Elemente dieser Denkweise. Gerade die Vorstellung der unabhängig von Menschen agierenden Technik bedient beispielsweise eine Medienlogik, die nach Dramatisierung sucht. Sie ist aber auch tief verwurzelt in einem alltagsweltlichen Denken, wenn Technik als Mittel der menschlichen Machtsteigerung gedacht wird. Des Weiteren zeigen sich technikdeterministische Denkweisen auch heute in der Rede zur Digitalisierung wieder, indem betont wird, Digitalisierung werde die Gesellschaft massiv verändern und transformieren, was man nicht verpassen dürfe. *Die Technik* führe, so war es auch historisch häufig zu lesen, zu einer neuen Gesellschaft, als sei sie monokausale Ursache des Wandels. Ähnlich wurde in den 1970er/80er Jahren über die Mikroelektronik gesprochen, aber auch die Rede von der Informationsgesellschaft spiegelt dies.

Aus historischer Perspektive sind die Diskurse um die Autonomie der Technik, wie sie bis in die 1970er Jahre hinein dominierten, aber nicht nur interessant, weil sie derartig noch immer alltagsweltliche Denk- und damit auch Verhaltensweisen prägen. Insbesondere das mit dem Technikdeterminismus verknüpfte anthropozentrische Konzept des Menschen als Gestalter und Beherrscher der Technik ist auch insofern für das Verständnis heutiger Diskurse relevant, als autonome Technik derzeit als Werkzeug, als Mittel interpretiert wird, mit dem Menschen zu mehr Autonomie gelangen könnten und damit die klassische Figur der Steigerung menschlicher Macht bedient wird. Dies wird im dritten Teil ausgeführt.

In den kulturkritischen Diskursen meinte Autonomie, zusammenfassend auf den Punkt gebracht: Macht. Macht, die Menschen nicht an Technik abgeben wollten. Der Diskurs war gleichermaßen Selbstvergewisserung wie Bekräftigung dieses menschlichen Anspruchs auf die zentrale Position in einer technischen Kultur. Eine ähnlich anthropozentrische Perspektive findet sich in den Diskursen um Automaten und um Automatisierung.

2. AUTOMATEN UND AUTOMATISIERUNG: HERR UND KNECHT UND DER ÜBERFLÜSSIGE MENSCH

Vom Altertum bis in die Neuzeit meinte der Begriff des Automaten eine von Menschen konstruierte Statue, die menschliche Bewegung oder Funktionen nachahmen konnte (vgl. Schnelle 2010). Im 17. Jahrhundert verglich René Descartes den menschlichen Körper bekanntermaßen mit einem Automaten, betonte damit aber zugleich die Sonderstellung der Menschen, deren Geist und Seele sie von Maschinen und Tieren, die gleichfalls Automaten seien, unterscheide.

Wie der Begriff des Automaten ist auch jener der Automatisierung vor allem aber mit der Nachahmung oder Ersetzung menschlicher Tätigkeiten verbunden. Aristoteles sprach von Automaten als Werkzeugen, die Arbeit ohne menschliche Beteiligung, ‚von selber‘ verrichten: „[D]ann brauchten allerdings die Meister keine Gesellen und die Herren keine Knechte mehr“ (Aristoteles 1981: 4), so schlussfolgerte Aristoteles. Bereits in der Antike wurde mithin die Idee der Übertragung von Tätigkeiten auf Maschinen und der Ersetzung menschlicher Arbeit formuliert. Bereits hier findet sich das Motiv von Herr und Knecht, das sowohl den Diskurs um die Autonomie der Technik als auch um die Automatisierung stets grundierte.

Der Gedanke der Selbsttätigkeit von Maschinen, ihres Ablaufens *ohne menschliche Eingriffe*, etablierte sich, auch wenn er sich bereits in der Frühen Neuzeit findet, deutlich im Kontext der Industrialisierung im 19. Jahrhundert und wurde zu einem wirkmächtigen und selbstverständlichen Konzept der menschlichen Arbeitswelt, mit je unterschiedlichen Konnotationen, was die Position der Menschen anging.

Prominente und recht gegensätzliche Einschätzungen für das 19. Jahrhundert finden sich beispielsweise bei Andrew Ure und Karl Marx. Ure äußert sich begeistert über die Möglichkeit der „Automatisierung“ in der Textilindustrie, die eine billigere und präzisere Produktion ermögliche und die „irregularities“ menschlicher Arbeit beseitige. Er in-

terpretiert die automatisierte Technik als der menschlichen Hand überlegen (vgl. Ure 1835: 37 und 331).

Marx konstatierte erst einmal die Menschenlosigkeit eines „automatischen Systems“: „Sobald die Arbeitsmaschine alle zur Bearbeitung des Rohstoffs nötigen Bewegungen ohne menschliche Beihilfe verrichtet und nur noch menschlicher Nachhilfe bedarf, haben wir ein automatisches System der Maschinerie.“ (Marx 1970 [1890]: 403) Gleichermaßen sprach Charles Babbage im Kontext der Entwicklung von Rechenmaschinen von „Automaten“, die selbsttätig seien, sofern sie in der Lage seien, „allein durch den Antrieb einer Feder, eines herabhängenden Gewichts oder irgendeiner anderen konstanten Kraft zu ihren Resultaten zu kommen“ (Babbage 1996 [1864]: 73).

Während auch Babbage, wie Ure, darin einen Vorteil gegenüber dem fehlerhaften Menschen sah, verband Marx mit dem Einsatz der Maschine in der kapitalistischen Produktionsweise eine fundamentale Abwertung der menschlichen Arbeit. Der Mensch diene der Maschine und werde zu ihrem lebendigen Anhängsel, zum „Teil einer Teilmaschine“. Das „Talent“ des Arbeiters werde „verdrängt“, die „Virtuosität in [der] Führung [des Arbeitswerkzeugs]“ gehe „vom Arbeiter auf die Maschine über“ (Marx 1970 [1890]: 442): Der Arbeiter sei „geistig und leiblich zur Maschine herabgedrückt“, zum „Blödsinn“ und zur „Verkrüppelung“ verdammt (Marx/Engels 1968 [1844]: 513). Aber auch der Maschinenbauingenieur Franz Reauleaux konstatierte im 19. Jahrhundert „die auffallende Abnahme der Geschicklichkeit der Arbeiter. [...] [D]er Mensch aber, ihr Diener – grausige Ironie – sinkt auf die Stufe der Maschine herab“ (Reauleaux 1875: 519).

Im 20. Jahrhundert schließlich, insbesondere in der zweiten Hälfte, wurde Automatisierung im Kontext der Computerisierung zu einem selbstverständlichen Prinzip der Produktion und meinte nicht mehr nur, wie im 19. Jahrhundert, Mechanisierungsprozesse.

In den 1930er Jahren, noch mehr aber in den 1940er und 1950er Jahren, wurden in der amerikanischen Automobilindustrie neue Produktionsprinzipien entwickelt, die als *Detroit Automation* bezeichnet wurden. 1946 hatte Del Harder, Vizepräsident bei Ford, damit ein Fer-

tigungsverfahren bezeichnet, „bei dem das Werkstück ohne Mensch [in die Maschine] eingeführt, in ihr bewegt und zur nächsten Maschine transportiert wird“ (zitiert nach Dolezalek 1966: 217). Auch der Ingenieur Carl Dolezalek definierte Automatisierung kurz und bündig und zugleich legitimatorisch mit der Menschenlosigkeit von Prozessen: Der Mensch brauche weder „ständig noch in einem erzwungenen Rhythmus für den Ablauf tätig werden“ (ebd.). Dies meinte jedoch nicht nur, wie suggeriert, Entlastung, sondern evozierte Ängste vor der Ersetzung und dem Überflüssigwerden der Menschen.

Mit der Automatisierung von Kontrolle und Steuerung seit den 1950er Jahren verband sich erneut eine kontroverse Diskussion um die Position der Menschen. Die Einschätzungen changierten zwischen der Angst vor der Entmachtung der Menschen als Kontrolleur/-innen und geistige Wesen, da Technik nun einerseits in der Lage sei, mit der Übernahme von Steuerungstätigkeiten eigenständig geistige Arbeiten auszuführen. Auf der anderen Seite wurde gerade Automatisierung mit dem Versprechen der Aufwertung menschlicher Arbeit legitimiert. Menschen würden aufsteigen zu alles kontrollierenden Knöpfedrücker/-innen, die von körperlichen Arbeiten und Routinetätigkeiten entlastet seien (vgl. Heßler 2015).

Unverkennbar basieren die Einschätzungen erneut auf der Vorstellung von Herr und Knecht sowie auf der Angst vor Ersetzung der Menschen. Eine Kernfrage war, wie auch heute wieder, die nach der Notwendigkeit menschlicher Arbeit in automatisierten Systemen bzw. ihrem Überflüssigwerden. Die Rede von der menschenleeren Fabrik in den 1980er Jahren gab dem reichlich Nahrung.

Seit der zweiten Hälfte des 20. Jahrhunderts veränderte und verschob Automatisierung, wie zuvor die Mechanisierung, das Verhältnis von Menschen und Maschinen in der Arbeitswelt, allerdings nicht immer in der erwarteten Weise. So machten Automatisierungsversuche auch deutlich, was Maschinen (noch) nicht können und unterstrichen umso deutlicher die Bedeutung und Notwendigkeit menschlicher Fähigkeiten. Lisanne Bainbridge sprach von der Ironie der Automatisierung (vgl. Bainbridge 1987). Sie beschrieb damit die Beobachtung,

dass gerade in hochautomatisierten Systemen die Rolle der oft wenigen, jedoch verbleibenden Menschen umso wichtiger ist, sei es für die Fehleranalyse, die Reparatur, für bestimmte Fingerfertigkeiten oder das Treffen von Entscheidungen. Menschen kommt in Automatisierungsprozessen jedoch auch eine ‚Restfunktion‘ zu, die allerdings, ganz wie es Bainbridge beschreibt, unabdingbar für das Funktionieren ist.

Während die kulturkritische Debatte um ‚autonome Technik‘ auf die Einflusslosigkeit der Menschen auf technische Entwicklungen als Ganzes fokussierte, war der Begriff der Automatisierung vor allem mit der Idee der ‚Menschenlosigkeit‘ in Arbeitsprozessen verknüpft. Dabei ging und geht es zumeist um die Frage der Ersetzung sowie um die zu sichernde zentrale Position der Menschen. Beide Diskurse sind auffällig von einer Angst grundiert, die menschliche Position sei eine untergeordnete, marginalisierte oder die Menschen gar gänzlich überflüssig.

3. AUTONOME TECHNIK: AUTONOMIE DURCH ZAHLEN

Automatisierung bezeichnet Prozesse, deren Abläufe vorprogrammiert und nicht veränderbar, damit aber auch vorhersehbar sind. Steuerung, Regelung und Ablauf wurden von einem Programmierer/einer Programmiererin festgelegt, wie es auch bei der eingangs genannten Performance Tinguelys imitiert wurde. Bei der im heutigen Sprachgebrauch autonomen Technik, also bei Anwendungen der Künstlichen Intelligenz (KI), handelt es sich dagegen um selbstlernende und adaptive Systeme. Auf der Grundlage von Daten erzeugen Computer selbst die Algorithmen, um ein Problem eigenständig zu lösen. Das bedeutet, dass sie ein vorgegebenes Ziel selbständig erreichen können. Demnach *autonome* Technik kann flexibel auf Situationen reagieren, *Entscheidungen treffen* und sich anpassen. Die Lösungswege der Technik sind allerdings nicht prognostizierbar.

Der Begriff der Technik wird hier nun weniger im Sinne des Kollektivsingulars *die Technik* benutzt, wie es im kulturkritischen Diskurs

der Fall war. Vielmehr bezieht er sich auf konkrete, digitale Technik und Künstliche-Intelligenz-Anwendungen.

Der Begriff der Autonomie wiederum bezieht sich, wie schon bei Automatisierungsprozessen, auf die Abwesenheit menschlicher Eingriffe, auf das Funktionieren ohne menschliche Tätigkeiten, nun aber bei selbständiger Problemlösung und Zielerreichung, die die KI-Anwendungen gerade auszeichnet: „Ein System wird erst dann als autonom bezeichnet, wenn es ohne menschliche Steuerung oder detaillierte Programmierung ein vorgegebenes Ziel selbstständig und an die Situation angepasst erreichen kann.“ (Fachforum Autonome Systeme im Hightech-Forum 2017: 5) Hätte Tinguely ein Autonomes System entworfen, dessen Ziel es ist, nur sich selbst zu zerstören, hätte es – im Idealfall – angesichts des unvorhergesehenen Brandes, der Anderes außer dem System selbst gefährdet, die Feuerwehr gerufen.

Historisch betrachtet sind es meist viele kleine Automatisierungsschritte, die der Autonomie der Technik vorausgehen. Der Schritt zur *Autonomie* bedeutet gleichwohl eine entscheidende Zäsur im Mensch-Maschine-Verhältnis, mithin einen qualitativen Sprung in einer kontinuierlichen Entwicklung der Delegation von Tätigkeiten an Technik, was sich am autonomen Fahren veranschaulichen lässt. Bereits im Jahr 1914, als ein Autopilot für Flugzeuge präsentiert wurde, diskutierte man, ob auch Automobile selbst fahren könnten. Was zu dieser Zeit, in der Automatisierung technisch noch kaum möglich war und phantastisch schien, findet sich in den folgenden Dekaden immer wieder als Vision, wie Fabian Kröger herausgearbeitet hat (vgl. Kröger 2012). Seit den 1950er Jahren fahren Menschen mit Fahrerassistenzsystemen: dem Tempomat, weiter dem Anti-Blockier-System, Bremshilfen, dem Abstandshalter bis hin zu Einparkhilfen. Viele Tätigkeiten beim Autofahren sind automatisiert. So disruptiv das autonome Fahren auch erscheinen mag und so zweifellos es eine qualitativ neuartige (Weiter-)Entwicklung darstellt, so stellt es gleichwohl auch eine langfristige inkrementelle Entwicklung dar. Automatisierungen einzelner Funktionen werden weiterentwickelt bis – und das ist das Entscheidende an auto-

nomer Technik – das ganze System autonom, das heißt, ohne menschliche Eingriffe, flexibel, bei selbständiger Zielerreichung, funktioniert.

Ein Anwendungsbeispiel für Autonome Systeme ist das *Smart Home*. Unterschiedliche Funktionen und Tätigkeiten im Haushalt, wie beispielsweise Heizung, Stromversorgung und Beleuchtung, werden vernetzt. Vor allem zeichnen sich die Autonomen Systeme aber dadurch aus, dass sich die Technik selbstlernend an die Gewohnheiten der Bewohner/-innen anpasst und dementsprechend die Energieversorgung regelt. Autonome Systeme schreiben aber beispielsweise auch einfache Sportreportagen, werden im Personalwesen oder im Banken- und Versicherungswesen eingesetzt, wo sie Entscheidungen über Kreditfähigkeit oder Vertrauenswürdigkeit von Menschen treffen. Die Polizei experimentiert mit der Vorhersage von Kriminalität, z.B. Einbrüchen, oder mit der Vorhersage der Rückfallwahrscheinlichkeit von Strafgefangenen.²

Autonome Technik wird derzeit in einzelnen Bereichen angewendet und ist für Menschen bereits konkret erfahrbar. Sie ist Bestandteil alltäglicher Praktiken, die das Zusammenspiel von Menschen und Maschinen verändert und Fragen nach Verantwortung, Entscheidungskompetenz und Haftung evoziert. Im Folgenden wird auf die Semantik des Autonomiebegriffs und die Frage nach dem Mensch-Maschinen-Verhältnis fokussiert, insbesondere auf die Position der Menschen.

Autonomie und Künstliche Intelligenz

Betrachtet man vor diesem Hintergrund jene neue Semantik des Autonomiebegriffs, wie er im Kontext von Künstlicher Intelligenz erscheint, so wird deutlich, dass er eine Technik bezeichnet, die Handlungen oder Entscheidungen ohne menschlichen Eingriff vollzieht und flexibel auf ihre Umwelt reagieren kann. Das heißt, Autonomie der Technik bedeutet: Technik, die in einer offenen Umwelt komplexe

2 Katharina Zweig spricht von „algorithmischen Entscheidungsunterstützungssystemen“ (Zweig 2017: 19).

Probleme ohne Eingriffe von Menschen lösen kann. Die Rede von autonomer Technik geht auch hier stets mit den Begriffen eigenständig und selbständig einher. Anders als in den Diskursen der 1950er und 1960er Jahre ist dieses Eigenständige und Selbständige allerdings nicht gleichbedeutend mit der Vorstellung, Autonome Systeme seien unabhängig von kulturellen und sozialen Werten. Eigenständigkeit bezieht sich allein auf das Problemlösen und Zielerreichen. Dass Algorithmen stets mit Sozialem und Kulturellem vermischt sind, wurde in jüngster Zeit gerade am oben genannten Beispiel der Rückfallhäufigkeit von Kriminellen diskutiert, gleichermaßen in der Kreditvergabe. Auch der viel debattierte Chatbot Tay wird in diesem Zusammenhang als anschauliches Beispiel genannt.

Autonome Technik wird daher nicht als eigenständig im Sinne einer von Gesellschaft unabhängigen Technik wahrgenommen, wie es in den 1950er und 1960er Jahren diskutiert wurde, oder gar wertneutral, sondern immer auch als das Produkt sozialer und kultureller Prämissen, die sich in den Daten und Algorithmen wiederfinden. Der Autonomiebegriff bezieht sich mithin, um dies festzuhalten, lediglich auf die eigenständige Problemlösung ohne menschlichen Eingriff.

Diese Problemlösung der ‚autonomen Technik‘ handelt und entscheidet, das ist wichtig zu bemerken, in einer mathematisch-digitalen Logik, nämlich auf der Basis von Daten, mittels Berechnung und statistischer Verfahren und Algorithmen. Der Kern des Autonomiebegriffs der Aufklärung, der an Menschen gebunden war, bezeichnete allerdings eine Autonomie durch Vernunft, Reflexion und Urteilskraft. Hier liegt ein entscheidender Unterschied, der von der unreflektierten Verwendung des Begriffs Autonomie verwischt wird, wenn das Autonome der Technik auf das selbständige Zielerreichen reduziert wird. Der Unterschied zwischen einer Autonomie im Kant’schen Sinne, zu der Menschen fähig sind, und der technischen Autonomie der selbständigen Zielerreichung hat jedoch Konsequenzen für die Handlungen und Entscheidungen: Was bedeutet es, wenn Entscheidungen algorithmisch getroffen werden, mithin auf der Basis von Daten und Statistiken, in digitaler Logik und damit ohne menschliches Urteilsvermögen,

ohne Intuition, ohne Emotion? Computerentscheidungen folgen einer anderen Logik als menschliche Entscheidungen, obgleich die Begriffe der Autonomie, der Entscheidungen und Handlungen für beide Formen gleichermaßen genutzt werden.

Technik, insbesondere die Computertechnologie, hat schon immer Begriffe und Konzepte herausgefordert, die als den Menschen vorbehaltene Kompetenzen galten. Stets führte dies zu intensiven Debatten und verschob die Begrifflichkeiten.

Ein prominentes Beispiel ist die Debatte um den Begriff des Denkens und die Frage, ob Computer denken können. Alan Turing hatte in seinem einschlägigen Aufsatz die Frage *Can Machines Think?* gestellt und darauf eine pragmatische Antwort gegeben, indem er die Frage verschob auf die Frage der Unterscheidbarkeit maschinellen und menschlichen Denkens in konkreten Situationen (Turing 2007 [1950]). Die fundamentalen Fragen ‚Was ist denken?‘ und ‚Können Maschinen denken?‘ lehnte er ab. Gleichwohl entbrannte eine hitzige philosophische Debatte, in der der Begriff des Denkens aus menschlicher Sicht definiert und beispielsweise auf die Bedeutung von Erfahrung oder Intuition hingewiesen und Computern daher das Denken abgesprochen wurde.

Gleiches gilt für Begriffe wie Privatheit, Freundschaft oder Emotionen, deren Sinn sich einerseits durch Digitalisierung und soziale Medien zu verschieben schien, die andererseits in anthropozentrischem Sinne als ausschließliche Domäne der Menschen verteidigt wurden.

Die entscheidende Frage ist allerdings nicht, wie im nächsten Abschnitt argumentiert werden soll, ob Maschinen denken können, Gefühle haben oder autonom im menschlichen Sinne sind. Sie sind dies selbstverständlich nicht. Bis heute hat keine Maschine den Turing-Test bestanden. Es handelt sich um Konzepte, die menschlichem Vermögen entlehnt sind, auf Maschinen angewendet werden und dabei andere Implikationen, Logiken und andere Konsequenzen haben. Gleichwohl ist unbestritten, dass Maschinen inzwischen Tätigkeiten übernehmen können, die man als ausschließlich menschlich erachtet – jedoch und das ist die Pointe: auf andere Weise. Zu sprechen wäre daher conse-

quent von maschinellem Denken, maschinellem Fühlen und maschineller Autonomie.

Der entscheidende Unterschied zwischen maschineller Autonomie und menschlicher Autonomie ist, dass die maschinelle Autonomie nicht (jedenfalls bislang nicht) in der Lage ist, den gesetzten Zweck und die Lösungswege zu reflektieren, geschweige denn, dass sie selbst einen Zweck oder ein Ziel setzen könnte. Eine autonome Maschine kann, um das Eingangsbeispiel der sich selbstzerstörenden Maschine Tinguelys noch einmal aufzunehmen, nicht reflektieren, dass sie sich gerade selbst zerstört. Sie kann die Bedeutung dessen, was sie tut, in keiner Weise ermitteln.

Die Frage ist daher nicht die, ob Maschinen im menschlichen Sinne autonom sind, sondern welche Aufgaben man Maschinen überlassen möchte und welche Aufgaben den Menschen, wenn man die *Unterschiede* zwischen Menschen und Maschinen berücksichtigt. Was sollen Maschinen, die auf der Basis von Daten, Algorithmen, Statistik und Sensoren autonom agieren, übernehmen? Was Menschen, deren Autonomie auf Vernunft und Reflexion gründet? Wie wären Mensch-Maschinen-Relationen jeweils zu definieren?

Aufgaben- und Arbeitsteilungen zwischen Menschen und autonomer Technik

Die Bestimmung einer Aufgabenteilung zwischen Menschen und Maschinen ist keine Trivialität. Betrachtet man die Diskurse und Vorstellungen hierzu, so finden sich insbesondere hinsichtlich der Position der Menschen sehr unterschiedliche, teils widersprüchliche Argumentationsmuster. Gemeinsam ist ihnen jedoch stets die Betonung der Zentralität der menschlichen Position bzw. ihre Gefährdung.

Gerade der medien- und populärwissenschaftliche Diskurs um autonome Technik erinnert vielfach an die kulturkritischen Argumente der 1950er und 1960er Jahre, indem häufig recht undifferenziert davon die Rede ist, dass die Maschinen übernehmen, dass sie Menschen ersetzen und überflüssig machen würden. Hier zeigt sich die anthropo-

zentrische Angst vor Ersetzung und Ohnmacht der Menschen. Eine autonome Technik bedroht demnach die menschliche Machtposition und könnte Menschen beherrschen und unterjochen.

Im ingenieurwissenschaftlichen Fachdiskurs wird dagegen stets die zentrale Position der Menschen betont, die Ersetzung und das Überflüssigwerden von Menschen dagegen als unrealistisches Szenario behandelt. Autonome Technik wäre demnach – anders als es sich in den Ängsten der Kulturkritik äußerte – kein Gegensatz zu menschlicher Autonomie, es ginge nicht um einen Machtkampf zwischen Mensch und Technik. Vielmehr findet sich zum ersten die klassisch-traditionelle Argumentation, dass autonome Technik im Dienste der Menschen, gleichsam als Instrument, verwendet werde und als Werkzeug die menschliche Autonomie verstärken könne. Das *Smart Home* oder das autonome Fahren, so beispielsweise ein Argument, könne zu einer „selbstbestimmten Lebensführung“ (Fachforum Autonome Systeme im Hightech-Forum 2017: 4) beitragen, gerade im Falle gesundheitlicher Einschränkungen. Lernende Systeme passen sich dabei, so das Versprechen, den Menschen an – nicht umgekehrt. Autonome Technik handelt und entscheidet in dieser Logik zwar selbständig und ohne menschliche Eingriffe, sie ist jedoch nicht autonom im Sinne einer unkontrollierbaren Entwicklung. Vielmehr steht sie im Dienste und unter Kontrolle der Menschen, die damit ihre Fähigkeiten verstärken oder mehr Bequemlichkeit genießen können. Ein unbemerkter Diener, um den sich Menschen nicht zu kümmern brauchen: die positive Seite des Golem.

Zum zweiten findet sich, insbesondere in diesen Fachdiskursen, die Rede von der Partnerschaft, der Kollaboration, der Assistenz, was nicht immer von der Idee der ‚Indienstnahme‘ von Technik zu trennen ist. Denn autonome Technik unterstütze Menschen, nehme ihnen Arbeit ab, bereite vor, assistiere. Dies bezieht sich sowohl auf händische Arbeiten in der industriellen Produktion, in der Roboter und Menschen eng zusammenarbeiten, als auch auf die erwähnten Beispiele von Reportage schreibenden Systemen oder den Einsatz von Künstlicher Intelligenz im Versicherungswesen etc.

Hier teilen sich Maschinen und Menschen die Arbeit, wobei, so wird ohne Unterlass betont, Menschen stets die letzten Entscheidungen treffen sollen oder treffen können sollen. Auch innerhalb der Partnerschaft zwischen autonomer Technik und Mensch wird mithin die zentrale Position der Menschen betont und diskursiv gesichert. Häufig ist auch davon die Rede, dass Technik die Routinetätigkeiten oder lästige Arbeiten übernehme, eine Argumentation, die eine lange Tradition in der Geschichte der Arbeit seit dem 19. Jahrhundert aufruft und implizit oder auch explizit auf das Herr-Knecht-Denken rekurriert.

Zu bedenken ist hier jedoch, dass jede Delegation von Aufgaben immer auch mit einem möglichen Kontrollverlust einhergeht bzw. des Vertrauens bedarf. Die Delegation von Aufgaben an Menschen, z.B. Mitarbeiter/-innen, erfordert Vertrauen und Kontrolle. Die Delegation von Aufgaben an autonome Technik bedeutet, nach dem derzeitigen Stand, pures Vertrauen, indem es einen Kontrollverlust impliziert, da die Wege der Programmierung und die Art und Weise, wie die Autonomen Systeme zu Lösungen kommen, für Menschen nicht nachvollziehbar sind. Technik wird hier zu einer Blackbox, die auch nicht für Expert/-innen durchschaubar ist. Damit ist eine neue, fundamental andere Dimension erreicht, da eine Technik hier mithandelt und mitentscheidet, die *prinzipiell* nicht nachvollziehbar ist.

Der KI-Forscher Marvin Minsky thematisierte diese Unvorhersehbarkeit und Eigenständigkeit bereits in den 1960er Jahren, also lange bevor maschinelles Lernen auf Basis neuronaler Netze möglich war:

„His [A.L. Samuel's, M.H.] argument, based on the fact that reliable computers do only that which they are instructed to do, has a basic flaw; it does not follow that the programmer therefore has full knowledge of (and therefore full responsibility and credit for) what will ensue. For certainly the programmer may set up an evolutionary system whose limitations are for him unclear and possibly incomprehensible.“ (Minsky 1963: 447 Fußnote 38)

Zudem basieren Vorentscheidungen der Algorithmen, die eventuell nicht nachvollziehbar sind, wie oben ausgeführt, auf reiner Informati-

onsverarbeitung und Berechnung, auf Daten und statistischen Verfahren, nicht auf Reflexionsfähigkeit und Urteilsvermögen. Die Algorithmen wissen nicht, was sie tun. Es stellt sich daher umso dringlicher die bereits aufgeworfene Frage, für welche Tätigkeiten, Praktiken und Aufgaben dies adäquat und gesellschaftlich konsensual ist.

Historisch findet sich bereits seit den 1960er Jahren die Vorstellung, dass sich menschliche und maschinelle Kompetenzen ergänzen könnten, gerade aufgrund ihrer unterschiedlichen Kompetenzen.³ Dies ist jedoch auch deshalb zu einfach gedacht, da die Delegation von Aufgaben nicht nur die Frage der Kontrollmöglichkeit bzw. nach dem Vertrauen in autonome Technik aufruft. Delegation bedeutet darüber hinaus in der Regel eine Verschiebung von Kompetenzen. Was Menschen nicht praktizieren, und dies gilt insbesondere für geistige Prozesse, verlernen und vergessen sie. Gerade für das Fliegen wurde häufig debattiert, was automatisiertes Fliegen, also die Delegation des Flugvorgangs an das Flugzeug für die Kompetenzen der Pilot/-innen bedeutet.⁴

Indem Menschen eng mit Maschinen zusammenarbeiten, entstehen *MenschMaschinen*-Hybride, die gemeinsam handeln. Die Aufgaben können nicht von einem Part allein bewältigt werden, es ist ein enges Aufeinander-angewiesen-Sein, eine untrennbare Kopplung, in der Menschen unter Umständen die Handlungen und Entscheidungen ihres Maschinenpartners grundsätzlich nicht mehr nachvollziehen können und mit den Algorithmen als Blackbox gemeinsam handeln und entscheiden.

Die zweite zentrale Frage ist daher, welche Kompetenzen Menschen verlernen und vergessen können. Historisch ist ein steter Wandel menschlicher Kompetenzen sowie Verschiebungen zu beobachten, gerade im Zusammenspiel mit Technik. Die Liste der verschwundenen

3 Vgl. beispielsweise den Beitrag von John Licklider (1960), in dem dieser den Computer als Erweiterung des menschlichen Verstandes interpretiert.

4 Johannes Weyer (1997: 239-257) thematisiert dies insbesondere für Unfallsituationen, in denen der Verlust von Erfahrungen und Kompetenzen entscheidend sein könnte.

Berufe und Tätigkeiten ist lang. Es gibt keine Gasriecher mehr, keine Aufzugführer, keine Telefonistinnen, um nur einige Beispiele zu nennen. Die Frage ist gleichwohl stets aufs Neue zu diskutieren, insbesondere wenn *autonome* Technik Aufgaben übernimmt.

Während also der gegenwärtige Diskurs die Indienstnahme autonomer Technik für die Autonomie der Menschen sowie die hervorgehobene Position der Menschen auch in Mensch-Maschinen-Kollaborationen betont, entstehen tatsächlich hybride Gemeinschaften von Menschen und Maschinen, die von gegenseitigen Abhängigkeiten und Undurchschaubarkeiten geprägt sind.⁵ Eine autonome menschliche Position ist in diesen hybriden Verhältnissen nicht möglich, insofern Menschen die Reflexionsfähigkeit und das Urteilsvermögen hinsichtlich der maschinellen (Vor- und Mit-)Entscheidungen verlieren. Es handelt sich schon längst um eine konstitutive Vermischung, um ein gemeinsames Handeln und Entscheiden von Menschen und Maschinen. Die Frage, die sich angesichts dieser Diagnose stellt, ist schlichtweg, welchen Part Menschen und welchen Maschinen übernehmen sollen. Diese Parts von Menschen und Maschinen sind unterscheidbar, auch wenn sie aufeinander bezogen sind und ineinander übergehen. Menschen und Maschinen denken, entscheiden, handeln *grundsätzlich* unterschiedlich. Nach der Konsequenz dieser Unterschiede gilt es zu fragen.

5 Darauf hat die Akteur-Netzwerk-Theorie schon lange hingewiesen. Sie stellt allerdings nicht die Frage nach Unterschieden und Asymmetrien.

4. AUTONOMIE DES SYSTEMS HEUTE: DER MENSCH ALS KOMPONENTE DER *TECHNOSPHERE*

Während im Mediendiskurs, im wissenschaftlich-fachlichen Diskurs von Ingenieur/-innen, Informatiker/-innen oder auch im politischen Diskurs, die menschliche Position stets betont und gesichert wird, also anthropozentrische Positionen vertreten werden, etablierte sich ein neuer theoretischer Diskurs, der die Debatte um die menschliche Autonomie und autonome Technik gänzlich anders aufrollt. Gemeint ist hier nicht die Akteur-Netzwerk-Theorie, sondern Ansätze, die systemisch denken und die insbesondere die Anthropozän-Debatte mit der Frage der Technik verbinden. Thema sind komplexe Systeme, die als autonom bezeichnet werden. Es geht nicht um einzelne Technologien.

Im Unterschied zu den kulturkritischen Debatten sowie auch zum gegenwärtigen öffentlichen Diskurs wird hier autonome Technik (als System) im Kontext der Überwindung des Anthropozentrismus und der Entlarvung der nur vermeintlichen Souveränität der Menschen diskutiert und auch gefeiert. Die Illusion der Menschen, Souverän und Kontrolleur zu sein, sei hinfällig.

Bereits Ende der 1970er Jahre hatte der bereits erwähnte Politikwissenschaftler Langdon Winner über autonome Systeme reflektiert. Er grenzte sich dabei von kulturkritischen Positionen ab, verwies jedoch deutlich auf die Unkontrollierbarkeit komplexer Systeme. Er sprach vom „loss of agency in technological systems“ (Winner 1977: 295). Weiter schrieb er „[...] the possibility of directing technological systems clearly perceived, consciously chosen, widely shared aims becomes an increasingly dubious matter“ (ebd.: 295f.). Er sprach gar von der „pathetic fantasy“ (ebd.: 296) der Kontrollierbarkeit. Anders als Ellul und andere beklagte Winner jedoch nicht die Machtlosigkeit der Menschen, sondern thematisiert das Problem der „moral agency“ in komplexen Systemen (ebd.: 304).

Ähnlich argumentiert nun der Geologe Peter Haff mit seinem Konzept der *Technosphere* (vgl. Haff 2014). Er wendet sich dezidiert ge-

gen anthropozentrische Perspektiven. Die *Technosphere* beschreibt er als ein technologisch-basiertes System, in dem Menschen ein Teil sind, ein Teil, der vom System abhängig ist, ohne es als Ganzes beeinflussen zu können. Als Geologe beschreibt er Technik neben Luft, Wasser und Biologie als eine weitere Sphäre, als weitere Komponente des Erdsystems. Diese *Technosphere* sei quasi-autonom und sie präge die Bedingungen des menschlichen Lebens. Die Kerngedanken lesen sich nicht unähnlich zu Ausführungen wie sie sich bei Ellul finden, auf den sich Haff bezieht. Haff argumentiert gegen die „natural assumption that the technosphere is primarily a human-created and controlled system“ (ebd.: 127). Stattdessen arbeitet er heraus, dass

„[...] the workings of modern humanity are a product of a system that operates beyond our control and that imposes its own requirement on human behavior. The technosphere is a system for which humans are essential but, nonetheless subordinate parts. At shorthand we can say that the technosphere is autonomous.“ (Ebd.)

Das Vokabular ähnelt der Kulturkritik, indem die Unabhängigkeit von menschlichen Akteur/-innen sowie die Unkontrollierbarkeit des Systems betont werden.

Haff fasst diese autonome Dynamik in sechs Regeln zusammen. Dazu gehört beispielsweise „the rule of impotence, that most humans cannot significantly influence the behavior of large technological systems“ (ebd: 126) oder „the rule of control“ (ebd.), in der er ausführt, dass Menschen nicht ein System beeinflussen können, das „express a larger number of behaviors than he himself“ (ebd.).

Ähnlich wie Ellul betont er letztlich die Eigenlogik des Systems, das seinen eigenen Anforderungen folgt („Autonomy has its own necessities.“ [Ebd.: 127]) und das nur an bestimmten lokalen Stellen, je-

doch nicht als Ganzes von Menschen beeinflussbar ist.⁶ Das System *Technosphere* hat eine eigene Logik und ist auf der *Large-scale*-Ebene autonom. Anders als Ellul setzt Haff Menschen und das autonome System *Technosphere* aber nicht gegeneinander; auch handelt es sich nicht um eine kulturkritische Position, die den Kontroll- und Machtverlust der Menschen beklagt. Vielmehr geht es darum, nicht allein vom Menschen her zu denken, sondern vom System. Er spricht von einer „system-centric perspective“ (ebd.).

Daraus folgt keine Angst, sondern Verantwortung, indem der Anthropozentrismus und der Glaube an die menschliche Macht, an Steuer- und Kontrollierbarkeit als Illusion entlarvt wird. Menschen sind nur Komponenten eines Systems, das als Ganzes autonom ist; sie sind *Human Components* neben vielen anderen *Components*.

Diese Haltung fordert die Annahme der Gestaltbarkeit von Technik heraus. Die Gestaltbarkeit der Digitalisierungsprozesse, der Autonomen Systeme und ihrer Nutzung ist derzeit ein zentrales Thema. Haffs Ausführungen beziehen sich auf das System als Ganzes. In einzelnen Bereichen könnten Menschen steuern und Entwicklungen beeinflussen, so betont auch Haff. Dies betrifft beispielsweise Entscheidungen hinsichtlich des autonomen Fahrens, autonomer Drohnen oder von Assistenzsystemen in Fabriken. Jedoch entziehe sich der menschlichen Entscheidung, was dabei mit dem Gesamtsystem, auf der *Large-scale*-Ebene, geschieht.

Menschen müssen gestalten, doch sollten sie dabei nicht glauben, sie seien autonom und *bedienten* sich autonomer Technik. Vor allem stellt sich die Frage nach dem Zusammenwirken vieler einzelner autonomer Technologien – eine Frage, die noch gar nicht diskutiert wird.

6 Haff erwähnt Ellul und Winner und betont, dass er zur Diskussion um die Autonomie der Technik die Frage nach der „physical basis“ der Mensch-Technik-Beziehung beitragen möchte (Haff 2014: 127).

LITERATUR

- Aristoteles (1981): Politik, Hamburg: Felix Meiner.
- Babbage, Charles (1996 [1864]): „Die Difference Engine“, in: Bernhard Dotzler (Hg.), Babbages Rechen-Automate: Ausgewählte Schriften, Wien/New York: Springer, S. 73-92.
- Bainbridge, Lisanne (1987): „Ironies of Automation“, in: Jens Rasmussen/Keith Duncan/Jacques Leplat (Hg.), New Technology – and Human Error, Chichester: John Wiley & Son, S. 276-283.
- Dolezalek, Carl (1966): „Was ist Automation?“, in: Werkstatttechnik 56/5, S. 217.
- Ellul, Jacques (1964): The Technological Society, Toronto: Alfred Knopf.
- Fachforum Autonome Systeme im Hightech-Forum (Hg.) (2017): Fachforum Autonome Systeme. Chancen und Risiken für Wirtschaft, Wissenschaft und Gesellschaft, Abschlussbericht, Berlin: Fachforum Autonome Systeme im Hightech-Forum.
- Haff, Peter (2014): „Humans and Technology in the Anthropocene: Six Rules“, in: The Anthropocene Review 1/2, S. 126-136.
- Heßler, Martina (2015): „Die Ersetzung des Menschen? Die Debatte um das Mensch-Maschinen-Verhältnis im Automatisierungsdiskurs“, in: Zeitschrift für Technikgeschichte 82/2, S. 109-136.
- Kröger, Fabian (2012): „Fahrerlos und unfallfrei“, in: Uwe Fraunholz/Anke Woschek (Hg.), Technology Fiction. Technische Visionen und Utopien in der Hochmoderne, Bielefeld: transcript, S. 93-114.
- Licklider, John (1960): „Man-Computer Symbiosis“, in: IRE Transactions on Human Factors in Electronics, HFE-1, S. 4-11, <http://groups.csail.mit.edu/medg/people/psz/Licklider.html> vom 7.6.2018.
- MacKenzie, Donald/Wajcman, Judy (1985): „Introductory Essay and General Issues“, in: dies. (Hg.), The Social Shaping of Technology, Milton Keynes: Open University Press, S. 2-25.

- Marx, Karl/Engels, Friedrich (1968 [1844]): Ökonomisch-philosophische Manuskripte, in: Karl Marx/Friedrich Engels, Werke, Ergänzungsband 1, Berlin Ost: Dietz, S. 465-588.
- Marx, Karl/Engels, Friedrich (1970 [1890]): Werke – Das Kapital. Kritik der politischen Ökonomie, Bd. I: Der Produktionsprozeß des Kapitals, Berlin Ost: Dietz.
- Minsky, Marvin (1963): „Steps Toward Artificial Intelligence“, in: Edward A. Feigenbaum/Julian Feldmann (Hg.), Computers and Thought, New York: McGraw-Hill, S. 406-450.
- Pohlmann, Rosemarie (2010): „Autonomie“, in: Joachim Ritter/Karlfried Gründer/Gottfried Gabriel (Hg.), Historisches Wörterbuch der Philosophie, CD-Rom, Berlin: Schwab.
- Reauleaux, Franz (1875): Theoretische Kinematik. Grundzüge einer Theorie des Maschinenwesens, Braunschweig: F. Vieweg und Sohn.
- Ritter, Joachim/Gründer, Karlfried/Gabriel, Gottfried (Hg.) (2010): Historisches Wörterbuch der Philosophie, CD-Rom, Berlin: Schwab.
- Schelsky, Helmut (1961): Der Mensch in der wissenschaftlichen Zivilisation, Köln/Opladen: Westdeutscher Verlag.
- Schnelle, Helmut (2010): „Automat“, in: Joachim Ritter/Karlfried Gründer/Gottfried Gabriel (Hg.), Historisches Wörterbuch der Philosophie, CD-Rom, Berlin: Schwab.
- Spengler, Oswald (1931): Der Mensch und die Technik. Beitrag zu einer Philosophie des Lebens, München: Beck.
- Staudenmaier, John M. (1985): Reweaving the Human Fabric, Cambridge, Mass./London: MIT Press.
- Turing, Alan (2007 [1950]): „Computermaschinerie und Intelligenz“, in: Karin Bruns/Ramón Reichert (Hg.), Neue Medien. Texte zur digitalen Kultur und Kommunikation, Bielefeld: transcript, S. 37-64.
- Ure, Andrew (1835): The Philosophy of Manufactures, London: C. Knight.

- Völker, Klaus (1994): „Nachwort“, in: ders. (Hg.), *Künstliche Menschen. Über Golems, Homunculi, Androiden und lebende Statuen*, Frankfurt a.M.: Suhrkamp, S. 425-496.
- Weyer, Johannes (1997): „Die Risiken der Automationsarbeit. Mensch-Maschine-Interaktion und Störfallmanagement in hochautomatisierten Verkehrsflugzeugen“, in: *Zeitschrift für Soziologie* 26/4, S. 239-257.
- Winner, Langdon (1977): *Autonomous Technology. Technics-out-of-Control as a Theme in Political Thought*, Cambridge, Mass./London: MIT Press.
- Zweig, Katharina (2017): „Technischer Kommentar: Die Macht der Algorithmen“, in: *Autonome Systeme. Wie intelligente Maschinen uns verändern. Jahrestagung des Deutschen Ethikrates*, 21.7.2017, S. 19-22, https://www.ethikrat.org/fileadmin/PDF-Dateien/Veranstaltungen/2017-06-21-Jahrestagung-Autonome-Systeme_UF.pdf vom 7.7.2018.

Haben autonome Maschinen Verantwortung?

Christoph Hubig

Verantwortlichkeit ist keine manifeste Eigenschaft, sondern beruht auf Zuweisung durch Dritte und/oder durch Selbstzuweisung. In verbreiteten umgangssprachlichen Redeweisen finden sich oftmals undifferenzierte Zuweisungen von Verantwortlichkeit, wenn Einflussfaktoren, Determinanten, begünstigende oder unglückliche Umstände etc. gemeint sind. Das Subjekt der Zuschreibung wird dabei oftmals unter einem allgemeinen Kollektivsingular gefasst, wenn etwa davon die Rede ist, dass die Digitalisierung inzwischen in höherem Maße als die Globalisierung für den gesellschaftlichen Wandel ‚verantwortlich‘, die E-Mobilität für den Wegfall zigtausender Arbeitsplätze in der Getriebefertigung ‚verantwortlich‘, das Ausbildungssystem für Chancenungleichheit ‚verantwortlich‘ oder der Einsatz Autonomer Systeme für die menschenleere Fabrik ‚verantwortlich‘ seien. Bei näherer Betrachtung zeigt sich schnell, dass solche Diagnosen mit ihren markigen Titelworten nicht zielführend sind, solange nicht die komplexen Beziehungsgeflechte freigelegt sind, innerhalb derer sich Interessen durchsetzen, Weichenstellungen für das Eröffnen oder Verschließen von Handlungsräumen vorgenommen sowie Entscheidungen zur Realisierung bestimmter kausaler Effekte selbst stattfinden oder beeinflusst werden. Wenn in diesem Kontext der Frage nach der ‚Verantwortlich-

keit‘ Autonomer Systeme nachzugehen ist, geraten wir in ein Spannungsfeld, innerhalb dessen Relationen zwischen Handlungen (i.e.S.) natürlicher Subjekte, handlungsförmigen Vollzügen nicht-menschlicher Akteure unter delegierter Verantwortlichkeit (wobei die Verantwortlichkeit für die Delegation bei menschlichen Subjekten verbleibt), und schließlich einem Prozessieren der Systeme, die freigesetzt und in einem gewissen Sinne ‚autonom‘ sind (eigenständig Problemlösungen zu entwickeln und zu ‚entscheiden‘), zu verhandeln sind. Hierbei sind in vielerlei Hinsicht Differenzierungen anzubringen, denn der Teufel liegt – wie immer – im Detail. Nachfolgend sollen hierfür in einigen Schritten klassische Diskussionslinien aufgenommen und Vorschläge unterbreitet werden, die geeignet sind, die Rede von der ‚Verantwortlichkeit‘ für unser Problemfeld ‚griffiger‘ zu machen und dadurch die Voraussetzung zu schaffen, Strategien für den Umgang mit Konsequenzen unterschiedlicher Verantwortlichkeit zu erarbeiten.

1. DER FRAGEHORIZONT BEZÜGLICH VERANTWORTUNG

Wenn wir mit Blick auf das Agieren Autonomer Systeme der Frage nach der Zuweisbarkeit von Verantwortung nachgehen, sind zunächst sieben *Hinsichten* des Fragens zu unterscheiden und in einer Art Landkarte zu verorten, weil aus diesen Hinsichten eben nicht sieben *Typen* von Verantwortung resultieren, sondern diese erst dadurch zustande kommen, dass diese Hinsichten in unterschiedlichster Weise aufeinander beziehbar sind und sich erst dadurch die Spezifik der Verantwortungsfrage für Autonome Systeme herauszubilden vermag:

Erstens muss nach dem Subjekt (Wer?) der Verantwortlichkeit gefragt werden: Mögliche Kandidaten hierfür sind natürliche Individuen (z.B. als Fahrer/-in eines autonomen Fahrzeugs oder als Beteiligte an Interaktionen/Koaktionen mit einem Roboter), Individuen als Rollenträger mit Pflichten und Vollmachten (z.B. mit Aufsichts- und Kontrollfunktion) sowie Mensch-Maschine-Hybride mit geteilter Hand-

lungsträgerschaft bei der Realisierung bestimmter Handlungsschemata (siehe den Beitrag von Ingo Schulz-Schaeffer in diesem Band). Ebenfalls als Kandidaten infrage kommen Organisationen und Institutionen, die von unterschiedlich mandatierten Individuen (vgl. hierzu Hubig 2007: Kap. 7.4) vertreten werden (z.B. was die Erstellung von Systemarchitekturen, die Gestaltung der Trainingsbedingungen der Systeme und/oder die Regelwerke für die Mensch-System-Interaktion bis hin zur Haftung betrifft) sowie schließlich die neuerdings vom EU-Parlament vorgeschlagene artifizielle ‚elektronische Person‘ als mit einem gewissen Fonds versichertes Haftungskonstrukt (s.u.).

Zweitens muss nach dem Gegenstand (Wofür?) der Verantwortung gefragt werden: Hierbei kann es um singuläre Handlungen bzw. Handlungsfolgen in Gestalt direkt kausal hervorgerufener Effekte gehen (gelungene oder misslungene Aktionen), die Übernahme oder die Delegation von Rollen in der Mensch-System-Interaktion, den Aufbau, die Fortschreibung oder Zerstörung von Strukturen möglichen Agierens, den Erhalt oder die Missachtung von Gütern (z.B. ökologischer, ökonomischer oder sozialer Besitzstände), die Gestaltung von Bedingungen systemischen Lernens in Verbindung mit der Eröffnung bestimmter Freiräume des Agierens etc.

Drittens ist nach der Domäne der Verantwortungszuweisung zu fragen: Handelt es sich um moralische, rechtliche, ökonomische oder politische Verantwortung (letztere z.B. auch dann, wenn individuell-persönliches Verschulden nicht vorliegt)?

Viertens ist nach (möglicherweise konfligierenden) Maßstäben zu fragen: Handelt es sich um Werte, Normen, Prinzipien der Moral (als in einem bestimmten Kontext subjektiv anerkannter Orientierungsinstanzen) oder der Sittlichkeit (als in ihrer Rechtfertigbarkeit objektiv/ intersubjektiv nachvollziehbarer normativer Instanzen), oder handelt es sich um Rechtszustände mit ihren Regelwerken, die sich aus der Rechtsetzung, nachgeordneten Regelungen durch die Exekutive oder autorisierte Verbände (z.B. der Verein Deutscher Ingenieure [VDI], der Verband der Elektrotechnik Elektronik Informationstechnik e.V.

[VDE] oder das Deutsche Institut für Normung [DIN]) oder dem Richterrecht speisen?

Fünftens ist nach der Instanz (Wovor?) der Verantwortlichkeit zu fragen. Hier kommen das eigene Selbst mit seinen Ansprüchen (z.B. in Gestalt des ‚Gewissens‘), betroffene Dritte mit jeweils eigenen Ansprüchen, moralische Autoritäten mit ihren Lob- und Schuldzuweisungen, die Jurisdiktion, Kontrollgremien/-behörden etc. infrage.

Sechstens ist zu fragen, für welchen Zeitpunkt bzw. welche Zeitspanne Verantwortlichkeit zuzuweisen ist: retrospektiv für Vorfälle oder Entwicklungen in der Vergangenheit (deren Folgen oder Folgebelasten ggf. zu kompensieren sind) oder prospektiv angesichts möglicher Entwicklungen in der Zukunft, für die Vorsorgepflichten zu übernehmen sind (z.B. was den Umgang mit *Chancen- und Risikopotenzialen* betrifft).

Und schließlich ist *siebtens* der Raumbezug zu thematisieren: Handelt es sich um lokal eingrenzbare Phänomene (für die sich möglicherweise die Verantwortungsfrage leichter beantworten lässt) oder geht es um regionale bis a limine globale Effekte, die unter mannigfachen, multiplen und/oder synergetisch wirkenden Faktoren zustande kommen, deren konkrete Relevanz und Proportionierung schwieriger zu analysieren ist?

Auf dieser ‚Landkarte‘ sind diejenigen möglichen Verbindungen zwischen den ‚Orten‘ des Fragens mit Blick auf das Agieren ‚autonomer Maschinen‘ zu eruieren. Insbesondere wäre zu fragen, ob angesichts der Komplexität der Problemlage Strategien der Vereinfachung der Verantwortungszuweisung auch und gerade in praktischer Hinsicht rechtfertigbar erscheinen (siehe Kapitel 6).

2. KRITERIEN DER VERANTWORTLICHKEIT

Fragen wir zunächst in einem zweiten Schritt nach einer einfachen Ausgangsposition, von der aus Kriterien der Verantwortlichkeit entsprechend unseren Intuitionen ersichtlich werden, um dann weiter zu

fragen, inwieweit mit Blick auf autonome Maschinen hier Modifizierungen anzubringen sind. Für individuelle natürliche Subjekte gilt wohl als *erstes Kriterium* Autonomie als Fähigkeit, unter selbst (*autos*) gesetzten und/oder anerkannten Regeln (*nomoi*) zu handeln, das heißt Mittel zur Realisierung von Zwecken unter Zielen der Lebensführung einzusetzen. Dies setzt (negative) Freiheit von externer Determination (Freiheit von ...) und (positive) Freiheit der Verfügung über einen Optionenraum für Handlungsalternativen (Freiheit zu ...) voraus. Wäre beides jeweils nicht gegeben, würden wir Verantwortlichkeit absprechen oder zumindest entscheidend relativieren.

Zweitens gilt als Kriterium (welches mit dem ersten einhergeht), dass das Subjekt ‚intentionale Zustände‘ aufweist, das heißt in einem bewussten Weltbezug steht, innerhalb dessen es im Rahmen seines theoretischen und praktischen Wissens Gegenstände und Zustände identifiziert und qualifiziert, z.B. als wünschenswert oder zu vermeiden, als herzustellen oder zu beseitigen, als vorzuziehen oder zu vernachlässigen, als zu verändern oder in Kauf zu nehmen (Näheres hierzu in Kapitel 4).

Drittens muss das Subjekt als Autor seiner Handlungen im Sinne kausaler Urheberschaft erachtet werden können. Die Frage, ob dieses Kriterium jeweils erfüllt ist, führt uns in verzwickte Problemlagen bis hin in den juristischen Bereich, denn mit der bloßen Unterscheidung zwischen hinreichenden Bedingungen (als Ursachen) und notwendigen Bedingungen (ohne deren Vorliegen hinreichende Ursächlichkeit nicht zustande gekommen wäre) ist das Problem keineswegs erledigt. Es können nämlich Sachlagen entstehen, unter denen die Realisierung einer notwendigen Bedingung genau diejenige (einzige) Bedingung ausmacht, ohne die die vordergründig ‚hinreichende‘ Bedingung als solche gar nicht hätte wirksam werden können. Die Unterscheidung verschwimmt also und erfordert sorgfältiges Beurteilen und Abwägen, weil im Bereich natürlichen Handelns und sozialer Interaktion oftmals die klassischen Strategien naturwissenschaftlich-technischen Experimentierens mit seinen Kausalitätstests durch Weglassen oder Parametervariation nicht greifen. Dies schreibt sich – wie wir sehen werden –

in das Feld maschineller ‚Verantwortlichkeit‘ bei der Zuweisung kausaler Autorschaft fort.

Viertes Kriterium ist die Unterstellung einer Fähigkeit zur Kontrolle des eigenen Handelns/Agierens sowohl bezüglich der Anpassung, Korrektur, Modifizierung oder Aufgabe der Aktualisierung einer ursprünglich intendierten Handlung (z.B. angesichts veränderter Handlungsumstände) als auch und gerade (und damit einhergehend) die Kontrolle über Handlungsalternativen in ihrer parallel laufenden Entwicklung, und zwar mit Blick auf eine sinnvolle Änderung des Handlungsschemas oder Veränderungen bei der Priorisierung von Zwecksetzungen (z.B. angesichts von Opportunitätskosten). Ein solches kontinuierliches Lernen und Disponieren während des Handelns unterscheidet sich von einem Aktionstyp, der bloß als Umsetzung von vorher Gelerntem, Antrainiertem und Eingetübtem zu begreifen ist.

Schließlich gilt als *fünftes* Kriterium die Fähigkeit, das eigene Handeln unter normativen Gesichtspunkten zu rechtfertigen. Also die Frage ‚Warum hast du das getan?‘ nicht bloß bezüglich der gesetzten Mittel und Zwecke und entsprechender Funktionalitäten, sondern auch mit Blick auf die Rechtfertigungsfähigkeit dieser Setzungen, also im Zuge einer normativen Reflexion zu beantworten.

Wenn eine solche Verantwortlichkeit gegeben ist, können wir von einer ‚starken‘ (allein intrinsischen) Verantwortlichkeit sprechen. Diese ist für das Agieren von Mensch-Maschine-Hybriden oder gar für autonomes Agieren von Maschinen nicht zu unterstellen. In diesem Feld kommt eine ‚schwache‘ Verantwortlichkeit infrage, da partiell im Modus externer Steuerung und Regelung auf das Entscheiden und die Entscheidungsstrategien Einfluss genommen und lediglich für die entstehenden Freiräume ‚Verantwortlichkeit‘ delegiert wird. Von diesen Arten der Einflussnahme soll gleich die Rede sein; mit der Delegation von Verantwortlichkeit werden die Delegatoren natürlich nicht von einer Verantwortlichkeit für eben diese Delegation entlastet. Schließlich könnte noch von einer dritten Art von Verantwortlichkeit, einer ‚Als-ob‘-/simulierten Verantwortlichkeit für Maschinen gesprochen werden. Gemeint wäre damit ein moralanaloges Verhalten, wie es in den soge-

nannten Roboterethiken verhandelt wird. Eine äußere Ähnlichkeit reicht jedoch nicht für eine Verantwortungszuweisung, denn diese wäre dann auch allenfalls im Rahmen einer simulierten Zuweisung möglich, die an die Stelle expliziter Delegation schwacher Verantwortung in Gestalt funktionaler Vorgaben und qua Gestaltung der Trainingsbedingungen der Systeme tritt. Wir gehen dann (z.B. in juristischer Absicht/Haftung) mit den Systemen um, *als ob* sie verantwortlich wären (siehe Kapitel 6).

3. TYPEN VON AUTONOMIE

Zunächst lassen sich drei Typen von Autonomie unterscheiden, wobei der Übergang zwischen den ersten beiden Typen fließend ist (hierzu ausführlich Hubig 2015: Kap. 3.1.1; die Unterscheidung wurde vielerorts aufgegriffen, exemplarisch Schilling et al. 2016: 338-344):

a) *Operative Autonomie*: Sie findet sich in Systemen, deren Agieren Freiheitsgrade in der Wahl der *Mittel* nach Maßgabe von Effizienz und Effektivität aufweist. Zwar ist diese operative Autonomie dahingehend determiniert, dass die Zwecke des Agierens und der Spielraum/die Optionen der Wahl und der Adaption von Mitteln vorgegeben sind; die Maschinen operieren aber nicht *deterministisch* wie Automaten, deren Prozessieren auf entsprechenden Algorithmen fixiert ist. Smarte Assistenzsysteme jeglicher Art fallen unter jene Rubrik, sofern sie im Zuge von Lernprozessen ihr Verhalten in Anpassung an Umweltbedingungen und/oder Nutzerprofile optimieren. Ein Tempomat ist nicht operativ autonom, wohl aber elaboriertere Systeme, welche z.B. das Lenk- und Bremsverhalten unterstützen, Geschwindigkeiten oder generell das Fahrverhalten in Abhängigkeit von Umweltbedingungen und Vigilanzschwellen regulieren, unter vorgegebenen funktionalen Erfordernissen Fahrtrouten auswählen etc. und sich hierbei unter anderem auf einen ‚Erfahrungsschatz‘ auf statistischer Basis (s.u.), auf Bilanzierung von Chancen und Risiken der Steuerungsprozesse nach vorgegebenen Modellen sowie auf Belastungslimits von Material und Mensch

(z.B. nach Maßgabe von Geschicklichkeit, Konzentration und Ausdauer) stützen. Analoges gilt für die Koaktion mit Robotern, z.B. in Bereichen von Fertigung, Service, Reparatur, Therapie und Pflege (zu den hierbei auftretenden Problemen siehe Kapitel 5). Rahmenbedingung für die Nutzung entsprechender Freiheitsgrade und eine entsprechende Wahl von Steuerungsprozessen ist eine funktionale oder normative Orientierung, deren Ergebnis seitens der Entwickler/-innen bzw. in Kooperation mit Nutzer/-innen an die Systeme delegierbar ist und deren Konkretisierung und Aktualisierung/Verwirklichung innerhalb der Systeme entsprechend simuliert wird, *als ob* rationale Subjekte hier entschieden hätten.

Darüber hinaus ist b) eine *strategische Autonomie* identifizierbar: Auch hier wird kontextsensitiv agiert, jedoch nicht bloß, was die Wahl der optimalen Mittel, also die Wahl von Steuerungsprozessen betrifft. Vielmehr erhöhen sich die Freiheitsgrade auf die Wahl von *Zwecken* unter vorgegebenen allgemeinen Zielen, und zwar nach Maßgabe von deren Verwirklichungschancen und -risiken. Die Festlegung der konkreten Zwecke betrifft hierbei deren Priorisierung, Reihenfolge des Abarbeitens, Intensität relativ zum Aufwand (Grenznutzen), Perfektionsgrad, Taktung und Dauer der Zweckverfolgung sowie Direktheit oder Umweg der Realisierung bis hin zum Abbruch von Aktionen bei wenig aussichtsreich erscheinender Zweckrealisierung. Eine solche strategische Autonomie kann z.B. einem intelligenten System zugesprochen werden, welches eine gesamte Fertigungsanlage – unter Berücksichtigung unter anderem des Stands der Halbzeuganlieferung, eines sich abzeichnendem Ausfalls von Funktionen auf den Fertigungsinseln, Belastungslimits der Arbeitenden, Konsequenzen ausbleibender Zweckrealisierung, Opportunitätskosten, Gefährdung mittelfristiger Nachhaltigkeit der Produktion im Spannungsfeld zu kurzfristiger Erhöhung der Gratifikationen höherstufig – steuert und regelt, kurz: Planungsprozesse realisiert. Eine solche Planungskompetenz unter vorgegebenen Zielen ist graduell, das heißt unter Festlegung höher- oder niedrigschwelliger Eingriffstiefe des Menschen delegierbar. Dies berührt jedoch nicht die *Rechtfertigung* der Typisierung von Kontexten,

das heißt die Unterstellung der Relevanz und Gewichtung von Parametern bei deren Modellierung, und es betrifft nicht die Rechtfertigung von Werthaltungen derjenigen, die jene Planungskompetenz delegieren und dabei die Werthaltungen der von jenen Planungsprozessen Betroffenen zu berücksichtigen haben.

Wenn diese Dimension berührt wird ist eine c) *moralische Autonomie* gefordert: Sie liegt in der Freiheit der *Anerkennung* von Prinzipien der Systemarchitektonik (für b)), wobei unter einer solchen Freiheit der Anerkennung die jeweilige Anerkennung des Selbst als Subjekt der Anerkennung durch sich selbst mitlaufen muss; solcherlei macht im eigentlichen Sinne die Verantwortungsübernahme aus, die nicht (wie bei a) und b)) delegierbar oder *simulierbar* ist.

Mit der Leitdifferenz zwischen dem *Erkennen* von Regeln (einschließlich des Standes ihrer Umsetzung und entsprechender Reaktion hierauf) und dem *Anerkennen* der Validität von Regeln ist die Grenze markiert, unter deren Aspekt die Frage nach einem ‚verantwortlichen Handeln‘ autonomer Maschinen eine erste Antwort erfahren kann: Die Selbstzuschreibung einer verantwortlichen Anerkennung von Prinzipien und Zielen, das Selbst-Einstehen für Werthaltungen, die in ihrer Gesamtheit eine Persönlichkeit in verantwortlicher Urheberschaft für die Ausrichtung an höchsten Zielen ausmacht, bleibt noch so intelligenten Systemen verwehrt.

4. BEWUSSTSEIN ALS VORAUSSETZUNG FÜR MORALITÄTSFÄHIGKEIT

Aus meiner Sicht spricht nichts dagegen, intelligenten Systemen (wie auch höheren tierischen Spezies) elementares Bewusstsein zuzuweisen. Ein solches umfasst die spezifische Aktivität, angesichts von Erfahrungen als Präsentationen diese in *Repräsentationen* zu überführen. Repräsentationen sind Identifizierungen und Deutungen von Präsentationen als ... (z.B. als anzustreben oder zu vermeiden, als mit diesen und jenen Qualitäten versehen, als disponibel oder indisponibel, als verur-

sacht durch und sich auswirkend auf etc.). Diese Kompetenz zur Repräsentation erlangen ‚intelligente‘ Systeme im Zuge von überwachtem, unüberwachtem und bestärkendem Lernen (vgl. hierzu einführend Alpaydin 2008). Beim überwachtem Lernen werden Daten verwendet, die von Personen hinsichtlich einer bestimmten Bedeutung gekennzeichnet wurden. Dadurch werden bestimmte Eigenschaften (*Features*) in den Sensordaten zu Indikatoren für einen bestimmten Deutungsinhalt. Der Deutungsrahmen wird dabei von einem gewählten Kategoriensystem vorgegeben (vgl. Kotsiantis 2007; Lake et al. 2016). Beim unüberwachten Lernen gibt es keine Labels; vielmehr werden die Vorschläge für eine Deutung vom System selbst gemacht. Sie basieren auf dem Erkennen von Mustern in den Daten auf Basis einer repräsentativen Menge von Beobachtungen. Beim bestärkenden Lernen werden im Zuge eines Trainings über die Verknüpfung von *Features* und Deutungsinhalt qua Trainingsdaten hinaus systemische Reaktionen optimiert. Hierdurch gewinnen die Präsentationen, die uns durch unsere Sinnesorgane oder Sensorinputs zugänglich werden, Zeichencharakter (auf der Basis bestimmter Träger), vermögen sich aus ihrer situativen Bedingtheit abzulösen und zum Gegenstand von Erinnerung und von Planung zu werden. Sie werden (‚propositionale‘) Gegenstände eben eines Bewusstseins. Über einen bloßen Charakter als Information hinaus gewinnen solche Gegenstände den Status einer ‚Kontextinformation‘, wenn sie relevant sind; und dies sind sie, wenn sie bei natürlichen Subjekten eine Rolle in deren intentionaler Ausrichtung bzw. bei Systemen eine Rolle für die Ausfüllung der Systemfunktionen (Instanziierung von Variablen) spielen. Dadurch werden Systeme „kontextsensitiv“ (Abowd et al. 1999: 304.; vgl. auch Chalmers/Maccoll 2003). Beim *überwachten* Lernen werden die Trainingsdaten gelabelt, das heißt mit einem Bedeutungsetikett versehen, und beim *unüberwachten* Lernen werden gelabelte Daten als Trainingsdaten verwendet. Für diese sucht das System selbsttätig qua Clustering nach Merkmalen, die eine Gruppe von Samples, die dieses Merkmal haben, von anderen unterscheidet, und damit selbständig Vorschläge zu einer Kategorisierung bzw. zu Ordnungsstrukturen für die Unterscheidung von Gruppen von

Samples entwickelt. Dagegen läuft das *bestärkende* Lernen auf ein strategisches Vorgehen hinaus, auf Basis eines ‚internen Belohnungsmechanismus‘ qua Belohnungsfunktion die Strategie finden zu lassen, die für möglichst viele äußere Umstände die bestmögliche Teilentscheidung hinsichtlich eines gewünschten Ziels trifft (vgl. Alpaydin 2008: 398). Je nach Lerntyp resultieren entsprechend unterschiedliche Verantwortlichkeitsprofile bezüglich derjenigen Subjekte, die in jeweils spezifischer Weise die Systeme trainieren oder diese sich selbst optimieren lassen.

Solche Systeme vermögen mithin ‚lernend‘ (ihr Gegenstandsfeld erweiternd, sich korrigierend, sich adaptierend) mit den Informationen umzugehen. Dabei kann dieser Umgang unter vielen Determinanten unterschiedlichster Art stehen (evolutionär gebildeten Filtern, vorgegebenen und entwickelten Strategien der Fusion und Verarbeitung von Sinnesreizen bzw. Sensorimpulsen, Berücksichtigung von Signifikanzschwellen, erlernten Konventionen der Zeichenverwendung und -zuordnung, antrainierten Präferenzen und Aversionen etc.). Unter einer solchen Determiniertheit kann jener Umgang jedoch durchaus – wie gesagt – in sich nicht-deterministisch prozessieren.

Die Repräsentationen – und dies ist für Künstliche Intelligenz relevant – können dabei in Grenzen die Systemzustände des Bewusstseins selbst umfassen; Bewusstseinsformen können ein Selbstmodell bilden, welches unter anderem ihren Informationsstand, ihre Dispositionen des Agierens, ihre Leistungen und Grenzen und in Abhängigkeit hiervon die Bezugsbereiche möglicher Aktionen erfasst und bilanziert. Aus logischen Gründen können Selbstmodelle hingegen niemals ein *vollständiges* und für jedes ihrer Elemente, einschließlich der jeweils vollzogenen Selbstrepräsentation, gültig entscheidbares Selbstmodell bilden: Denn würde die Selbstrepräsentation selbst Element des Selbstmodells sein, wäre sie Objekt der Selbstrepräsentation und nicht mehr diese selbst. Würde sie nicht Element des Selbstmodells sein, wäre dieses unvollständig. Solche Bewusstseinsformen einschließlich ihrer Grenzen können als kognitive Basis den Autonomietypen a) und b) zugeschrieben werden.

Ganz anders verhält es sich hingegen mit einem spezifischen Typ von Bewusstsein, welches Menschen sich als Selbstbewusstsein zu schreiben. Es handelt sich hier nicht um ein Bewusstsein *über* das Selbst (als eine bestimmte Repräsentation), sondern um ein Bewusstsein *vom* Subjektcharakter eines Selbst. Von diesem Selbst als Subjekt ‚kann man sich kein Bildnis machen‘, denn es ist ja die vorauszusetzende Instanz der Produktion von Repräsentationen/Bildnissen. Zu einem solchen Bewusstsein als Selbstbewusstsein kann man sich nur im Modus der *Anerkennung* eines Selbst *entscheiden*, indem man sich höherstufig als verantwortlich dafür anerkennt, Prozesse des Anerkennens von Regeln, Prinzipien, Zielen zu vollziehen und gemäß dieser Anerkennung zu handeln. (Man kann sich auch diese Anerkennung verweigern, indem man sich distanziert und darauf verweist, dass man irgendwie fremdbestimmt war – ‚Das war nicht ich!‘) Dies betrifft auch und gerade die Notwendigkeit, sich zu den zahlreichen Prozessen, die wir als uns selbst determinierende Prozesse (qua Sozialisation, neuronale Verfasstheit, Konsumgewohnheiten etc.) erkennen, in ein Verhältnis zu setzen: Sobald solche Prozesse nämlich identifiziert sind, stehen wir notwendigerweise in einem Verhältnis zu ihnen, welches unabhängig von der Möglichkeit einer faktischen Einflussnahme auf diese Prozesse uns zumindest in die Position bringt, diese Prozesse als solche zu billigen, denen wir uns fügen wollen, oder die wir in dem Sinne nicht anerkennen, dass wir alles ins Werk setzen, eine Einflussnahme auf diese Prozesse zu erreichen. Insofern ist dies eine, vernünftigen Wesen vorbehaltene, Form eines Selbstbewusstseins, zu der man sich entscheiden kann oder nicht (im letzteren Falle machen wir uns – eine wählbare Option – zum Tier, begeben uns in eine Abhängigkeit oder lassen es dabei sein). Diese anerkennungsbasierte Form des Selbstbewusstseins kann nicht in Systeme Künstlicher Intelligenz implementiert werden. So wäre durchaus vorstellbar, dass Systeme Künstlicher Intelligenz im Rahmen ihres Selbstmodells feststellen, dass ihnen bestimmte Informationen fehlen, Lernprozesse noch nicht abgeschlossen sind, bestimmte Repräsentationen sich noch nicht stabilisiert haben bzw. zu hohe Schwankungsbreiten aufweisen etc. und sie

daher bestimmte Fragen nicht beantworten (und dies auch explizit ausdrücken) oder bestimmte Aktionen nicht ausführen können. Nicht aber können sie sich dahingehend äußern, dass sie angesichts einer bestimmten als notwendig erachteten funktionalen Vorgabe durchaus die Gelegenheit gehabt hätten, sich das entsprechende Knowhow und die entsprechenden Dispositionen des Agierens anzueignen, dies aber nicht gewollt hätten mangels Anerkennung der zugrunde liegenden Prinzipien.

Angesichts der Leitdifferenz zwischen *Bewusstsein* und *Selbstbewusstsein* (welches ein bloßes Selbstmodell überschreitet) kann die Antwort auf die Eingangsfrage nun dahingehend weiter entwickelt werden, dass Systemen Künstlicher Intelligenz mangels Selbstbewusstsein in diesem emphatischen Sinne die basale Voraussetzung für eine (nicht simulierte) Verantwortlichkeit und erst recht für ihre ethische Reflexion abzusprechen ist.

5. INTERAKTION MIT INTELLIGENTEN SYSTEMEN/ROBOTERN UND DIE HYBRIDISIERUNG VON AKTEUREN

Wenn es um den Umgang mit intelligenten Systemen geht, wird landläufig von Interaktion gesprochen. Keineswegs soll hier dieser Sprachgebrauch pauschal verurteilt werden, denn Begriffe unterliegen nicht gewissermaßen einem ‚Markenschutz‘. Gleichwohl ist anzumerken, dass im Rahmen dieses Sprachgebrauchs einige Pointen verloren gehen, die eine präzisere Begriffsverwendung mit sich führen. Das lässt sich insbesondere geltend machen mit Blick auf sogenannte moralische Dilemmata, in denen sich künstliche intelligente Systeme angeblich verfangen können. Gängige Beispiele hierfür sind das Verhalten autonomer Automobile im unvermeidlichen Kollisionsfall oder der Einsatz intelligenter Kampfdrohnen.

Um die Sicht auf die Problemlage zu schärfen, ist es sinnvoll, zunächst auf die engere Fassung des Interaktionskonzepts zu verweisen,

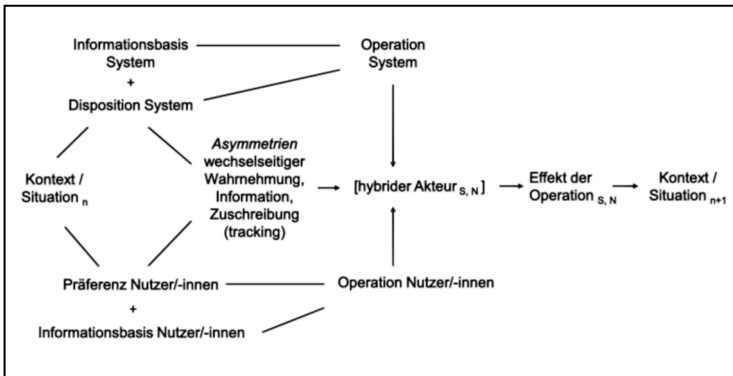
wie sie im Rahmen der Soziologie und der allgemeinen Handlungstheorie etabliert ist: Versteht man dort unter ‚Aktion‘ einen Prozess, der mit der *Erwartung* über die Realisierung eines Zwecks verbunden und durch diese motiviert ist, so bezeichnet ‚Interaktion‘ eine Aktionsform, die auf Aktionen anderer gerichtet ist und sich mithin an Erwartungen über die Erwartungen anderer, sogenannte *Erwartungserwartungen*, orientiert (vgl. Luhmann 1984: 412; Weber 1968: 441). Diese Erwartungserwartungen werden im Idealfall ständig einem kommunikativen Abgleich unterzogen (sodass in einem solchen Idealfall die Erwartungen der Entwickler/-innen über die Erwartungen der Nutzer/-innen und die Erwartungen dieser über die Erwartungen der Entwickler/-innen harmonieren). Solcherlei setzt entsprechende höherstufige elaborierte Kommunikationsprozesse voraus.

Für den Umgang mit intelligenten Systemen (einschließlich Robotern) sind solche höherstufigen Kommunikationsprozesse nicht in Sicht. Vielmehr findet die wechselseitige Bildung von Erwartungen bei ihren Trägern, Mensch und Roboter, *unabhängig* statt, und das Zusammenwirken menschlicher und künstlicher Systeme basiert auf unabhängiger und jeweils unterschiedlicher Art der Erwartungsbildung (vgl. Braun-Thürmann 2002: 15, 145). Die Systeme verfolgen qua *Tracking* die Aktionen menschlicher Nutzer/-innen und bilden hierauf basierend Muster bezüglich der eigenen Erwartungen, der Wertung der jeweiligen Elemente und Komponenten der Erwartung und der Wertung der Aktualisierung dieser Erwartungen in konkreten Aktionen (vgl. Fink 2009: 11; Krummheuer 2010: 105). Die natürlichen Interaktionspartner werden auf Basis ihrer Gewohnheiten und Handlungsrou-tinen auf Nutzerstereotype und Adressatenprofile reduziert (vgl. hierzu Hubig 2015: 138). Hieraus resultieren seitens der Systeme bestimmte *Dispositionen* des Agierens, die in den einzelnen Schritten ihres Zustandekommens für die Nutzer/-innen und in weiten Bereichen auch für die Entwickler/-innen intransparent bleiben. Lediglich in die Gestaltung der *Trainingsbedingungen* der Systeme bei ihrer Musterbildung und in die *Ergebnisse* des Agierens kann Einsicht genommen und über sie bilanziert werden. Erst recht kann sich die Musterbildung der

Systeme von konkreten Interaktionsprozessen lösen, wenn zusätzlich die Übernahme bestimmter Muster aus einschlägigen Kontexten des *World Wide Web* die Bildung von Dispositionen in den Systemen zusätzlich beeinflusst. Umgekehrt haben die Nutzer/-innen nicht mehr die Möglichkeit, ihre Erwartungserwartungen an die Entwickler/-innen der Systemarchitektur angesichts des Verhaltens der Systeme zu verifizieren, zu korrigieren, fortzuschreiben, weil das Feld der Determinanten der Bildung systemischer Dispositionen unüberschaubar und intransparent ist (vgl. Feiner 2002).

Daher wird neuerdings zur Charakterisierung dieser Vollzüge eher der Begriff ‚Koaktion‘ eingesetzt, und es ist die Rede von ‚hybriden Akteuren‘, wodurch signalisiert werden soll, dass zwei Agenten, die in ihrer Bildung von Mustern und Dispositionen auf der einen Seite und von Erwartungserwartungen auf der anderen Seite geleitet sind, ohne direkte Kommunikation, gleichsam arbeitsteilig, Handlungseffekte zeitigen (vgl. Rammert/Schulz-Schaeffer 2002: 11-64; Weyer/Fink 2011: 43).

Abbildung 1: Koaktion mit Autonomen Systemen



Quelle: eigene Darstellung

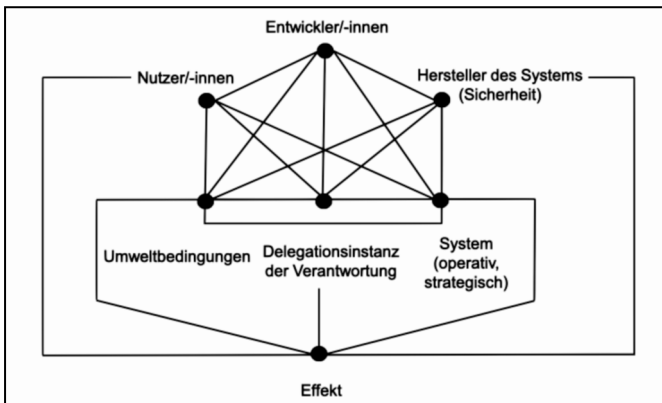
Sofern sichergestellt werden kann, dass die Systemdispositionen funktional äquivalent sind mit den Erwartungen der Nutzer/-innen, kann im Rahmen solcher Koaktionen eine gelingende Interaktion zumindest simuliert werden (z.B. beim Einsatz von Kuschelrobotern für Demenzerkrankte). Die Koaktion gerät jedoch an ihre Grenzen, sofern menschliche Akteur/-innen intrinsisch, das heißt ohne extern beobachtbare und registrierbare Stimuli, im Zuge von Abwägungen ihrer Erwartungen und Änderung ihrer Werthaltungen die Realisierung von Zwecken nicht zulassen oder verändern oder gewissermaßen die intelligenten Systeme ‚auszutricksen‘ suchen. Ein Beispiel hierfür wäre ein nicht auszuschließender Umgang mit autonomen Automobilen derart, dass aufgrund von deren Disposition, bei unvermeidlichen Kollisionen Personenschäden zu minimieren, Anreizsysteme für die Nutzung möglichst schwach gesicherter Automobile entstehen (weil die stärker geschützten bevorzugt angesteuert werden) oder bei vollbesetzten Fahrzeugen abwegige Lizenzen zum nicht-regelkonformen Verkehrsverhalten in Anspruch genommen werden (weil die Verursacher im Kollisionsfall eher verschont bleiben). Auch können zum Zwecke der Entledigung von Eigenverantwortung den Systemdispositionen allzu hohe Entscheidungskompetenzen zugewiesen werden, wenn es – wie im Falle von Kampfdrohnen – darum geht, potenzielle Kombattanten zu identifizieren und Fehlertoleranzen bei der Unterscheidung von Zivilpersonen und militärischem Personal in Kauf genommen werden.

Auch hier zeigen sich unüberbrückbare Leitdifferenzen, wenn auf der Menschenseite qualitative Erfahrungen wie Freude, Bedauern, Reue, die bestimmten Werthaltungen geschuldet sind, auf quantitativ-statistische Erfahrungen der Systeme (basierend auf ihrem jeweiligen Training) treffen, und umgekehrt deren individuell adaptive, lernende Musterbildung auf die Etablierung menschlicher Verhaltensmuster trifft, die sozialem Lernen, Rechtfertigungskulturen und moralischen Diskursen geschuldet sind.

6. VERANTWORTUNGSNETZWERKE

Angesichts der Hybridisierung des Menschen durch autonome Maschinen – wie in der Mensch-Roboter-Koaktion – (und umgekehrt einer ‚Hybridisierung‘ der Maschinen durch ihre Einbindung in Koaktionen mit Menschen, die die Maschinen nicht mehr bloß bedienen, sondern Instanz ihrer Lernprozesse sind) wäre zunächst kompromisslos auf Transparenz zu setzen, was die Trainingsbedingungen der Systeme, die Architektur der Herausbildung ihrer Dispositionen und die Bilanzierung der Aktionseffekte betrifft. Damit müsste eine adäquate Herausbildung von ‚Verantwortungsnetzwerken‘ einhergehen, innerhalb derer die Delegationsprozesse für Autonomie im Sinne von a) und b), die Möglichkeiten einer Verantwortungszuweisung in Haftungsfällen und die Prozesse der Bildung von Erwartungen bzw. Systemdispositionen der Beteiligten so verhandelt werden, dass Verantwortungsdiskurse *über* autonome Maschinen präzise und sachangemessen geführt und Voraussetzungen für einschlägige Verrechtlichungsprozesse erbracht werden. Solche Forderungen sehen sich freilich mit einer überaus hohen Komplexität der Sachlage konfrontiert.

Abbildung 2: Verantwortungsnetzwerk



Quelle: eigene Darstellung

Die Schwierigkeit gründet darin, dass in einem rekonstruierbaren Netz von untereinander wechselwirkenden Einflussgrößen für das Zustandekommen eines Effekts nicht auf einfachem Wege Verantwortlichkeiten in ihren jeweiligen Abhängigkeiten den Wirkungsrelationen zugeordnet werden können. Dies betrifft insbesondere Versuche einer Typisierung, Unterscheidung und einer jeweiligen Zuordnung von menschlicher (intentionaler), maschineller (technischer) Verantwortlichkeit oder aus ‚natürlichen‘ Umständen (Verfasstheit der Handlungsumgebung, des Materials und weiterer, nicht disponibler Gegebenheiten bis hin zu ‚Zufällen‘) resultierender Maßgeblichkeit. Denn bei allen Akteuren als Knoten in diesem Netzwerk finden wir eine Hybridisierung, in der intentionale, technische und natürliche Verfasstheiten sich überlagern und wechselwirken. Entsprechend können wir bestimmte Effekte, Teileffekte oder Seiten solcher Effekte nicht einfach auf das Konto intentionaler Verantwortung, technischer (delegierter) Verantwortung oder ‚der Natur‘ verbuchen. Der Soziologe Bruno Latour hat daher im Rahmen seiner Akteur-Netzwerk-Theorie (ANT) gefordert, die Rede von ‚der‘ menschlichen Subjektivität/Intentionalität, ‚der‘ Technik und ‚der‘ Natur zu verabschieden und stattdessen von ‚intentionalen Fragmenten‘, ‚technischen Fragmenten‘ und ‚natürlichen Fragmenten‘ zu sprechen, die sich nach Maßgabe der Relationen, in denen sie untereinander stehen, allererst entwickeln und aktualisieren (vgl. Latour 2006a, 2006b; hierzu ausführlich Hubig 2015: 88-95). So bilden wir unsere Absichten, Erwartungen und Präferenzen (insbesondere was deren Hierarchisierungen und Priorisierungen je nach Situation betrifft), einschließlich der entsprechenden Entscheidungen unter dem Einfluss der entsprechenden technischen Handlungsumgebung sowie den Artefakten samt ihrer materialen Verfasstheit mit ihren förderlichen und widerständigen Eigenschaften. Die technischen Komponenten wiederum verbleiben im Zustand toter Materialität, sofern sie nicht in diejenigen Handlungsprozesse eingebunden sind, in deren Lichte sie Funktionalitäten und ihre weiteren Eigenschaften anzunehmen vermögen. Und ‚Natur‘ gewinnt ihren Status allererst in ihrer Relation zu uns und zu möglichen technischen Funktionalitäten einschließlich ihrer inneren

Organisation nach Maßgabe technisch induzierter Modellierung und der Qualitätszuschreibung im Lichte weiterer (bis hin zu ästhetischer) intentionaler Konzeptualisierung. In allen Instanzen des Netzwerks können wir solche Relationen antreffen, wenn wir die Knoten nicht als ‚Blackboxes‘ annehmen, sondern ihre eigene interne Vernetzung freilegen. Diese Problemlage potenziert sich natürlich, wenn wir ein Netz in seiner Gesamtheit betrachten und die im Netz gezeitigten Effekte eben nicht mehr einfach auf Menschseite, Technikseite oder Naturseite verbuchen können. So hat Latour einschlägige Effekte von der Nutzung von Schlüsseln und Waffen über technische Maßnahmen der Verkehrsberuhigung, der Optimierung der Austernzucht u.v.a. mehr bis hin zu komplexen Netzen naturwissenschaftlicher Laborforschung, der Gestaltung von Verkehrsinfrastrukturen und ökologischen Projekten im Umgang mit dem Sahel-Syndrom akribisch untersucht und ausbuchstabiert. In diesem Zusammenhang wird regelmäßig deutlich, dass wir unter dem Problemdruck einer Orientierung unserer Aktionen in den Netzen zu Vereinfachungen, eben dem Blackboxing, gezwungen sind.

Solcherlei lässt sich auch angesichts unserer Problemlage beobachten und schließt den simplifizierenden Gebrauch des Titelwortes ‚Verantwortlichkeit‘ ein. Denn mit gewissen Einsichten, die sich aus einer differenzierteren Analyse ergeben, sind nicht zwangsläufig Fortschritte und Hilfen für den praktischen Umgang mit den Systemen verbunden. Im Bereich des Rechts findet dies seinen Niederschlag in der Diskussion, ob ein neues juristisches Konstrukt, die ‚e-Person‘ als Haftungsträger eingeführt werden soll (wie es vom EU-Parlament 2017 vorgeschlagen wird); sie könnte in zivilrechtlichen Verfahren verklagbarer Adressat für Haftungsansprüche werden und für die Schadenskompensation mit einem Fonds ausgestattet sein, der aus Ressourcen aller Akteure des Netzes rekrutiert würde, wobei die Proportionierung zu klären wäre. Eine andere Strategie läge in der Ausweitung der Gefährdungshaftung (die auch ohne Unterstellung eines Vorsatzes oder einer Fahrlässigkeit greift) über „bewegliche Sachen“ hinaus (ProdHaftG §2), wie sie etwa für Fahrzeughalter/-innen unabhängig

von der persönlichen Nutzung des Fahrzeugs besteht. Im Zuge dieser Ausweitung könnten gar auch Nutzer/-innen unter dem herkömmlichen Leitbild der Gefährdungshaftung, dass diejenigen die Haftungsrisiken tragen sollen, die Vorteile aus der Nutzung der Systeme ziehen, mit involviert werden (vgl. Günther 2016: 237-255). Auch eine Ausweitung der Fahrlässigkeit bzw. der Fahrlässigkeitskriterien unter Gesichtspunkten der Vorsorge und Sicherung kämen infrage, wobei dies jedoch in die Komplexität der Zuordnung von Zuständigkeiten zurückführen würde. Dies insgesamt ist aber nur eine Seite der Problematik.

Unter Verweis auf die Nicht-Delegierbarkeit moralischer Verantwortung folgt nämlich auf der anderen Seite, dass bezüglich der Großarchitektur der Systeme mit ihren sozialen und ökonomischen Konsequenzen auch klassische Strategien des Verweises von Zuständigkeiten an Politik, Wirtschaft, soziale Interessenvertretungen etc. an ihre Grenzen kommen und interessenfixierte Ziele in den Aushandlungsprozessen nicht mehr problemlos fortgeschrieben werden können (z.B. Erhalt von Arbeitsplätzen, Erhalt von Konkurrenzfähigkeit in bestehenden Marktstrukturen, Organisation von Bildungschancen angesichts veränderter Anforderungsprofile etc.). Die Herausforderungen an die Technikgestaltung und -nutzung sind so hoch, dass es nicht mehr zielführend erscheint, die Gesamtentwicklung dem Wechselspiel eines nutzenorientierten Interessenausgleichs zu überlassen, dessen Stakeholder jeweils als Einzelne und aus ihrer jeweiligen Binnenperspektive versuchen, das Bestmögliche für sich herauszuholen. Das betrifft sowohl die einzelnen Individuen als auch ihre Interessenverbände. Man muss nicht gleich Technokrat sein, um einen verstärkten Ausbau einer Ebene von Gremien zu fordern, die auf der Ebene einer Bilanzierung, Evaluation, Supervision und eines Monitoring der Entwicklung das Ziel zu verfolgen hätten, die politischen Kompetenzen sowohl der Legislative als auch derjenigen, die ihr die Macht verleihen, zu unterstützen und gleichzeitig zu entlasten: für erstere von der kurzfristigen Fixierung auf Machterhalt und -erweiterung, für letztere von der notwendigen Fixierung auf Erhalt und Erweiterung der Optionen ihrer individuellen Reproduktion auf Basis der problematisch gewordenen Arbeitsprozesse.

Für letztere würde dies einen hinreichenden Ausbau der Grundsicherung erforderlich machen, die eine gewisse Unabhängigkeit und mit Zeitressourcen versehene Anpassung an die Umwälzungen in der Arbeitswelt ermöglicht. Eine dialogische Abstimmung zwischen den Seiten der Entwicklung und denjenigen der Nutzung über Fragen der Architektur der Systeme sollte nicht mehr domänenspezifisch und quasi arbeitsteilig innerhalb der Wirtschaft, der Wissenschaft oder des Agierens der Verbände allein stattfinden, sondern in einem umfassenderen Rahmen, der von allen Beteiligten organisiert, getragen und gestaltet wird und bis zu einem gewissen Grade Appellations- und Legitimationsinstanz für politisches Entscheiden (mit) abgeben könnte. Hieraus würde für die Politik selbst auch eine Entlastung resultieren, und hierzu gibt es im Kleinen bereits Ansätze, z.B. die Chemie-Stiftung Sozialpartner-Akademie (CSSA) oder weitere einschlägige Institute gemeinsamer Trägerschaft von Unternehmensverbänden, Gewerkschaften und wissenschaftlichen Fachverbänden. Für diejenigen schließlich, die als Individuen in den Systemen arbeiten und mit smarten Systemen interagieren, sind die Optionen auszubauen, sich on demand bei Irritationen *parallel* zur Mensch-System-Kommunikation (vgl. hierzu Hubig 2007: 8.5; Wiegerling 2011 mit Report der Erträge des Sonderforschungsbereichs [SFB] 627 ‚Nexus‘) vertrauenswürdig und transparent auszutauschen nicht nur über die operativen Fragen, sondern auch über Hintergründe der strategischen Auslegung der Systeme, über Perspektiven ihrer Entwicklung bis hin zu Fragen von Ausbildungs- und Qualifikationsanfordernissen, erforderlicher Fortschreibung und Entwicklung von Kompetenzen und Situierung der eigenen Position bzw. Partizipation in diesem Prozess. Diesbezüglich bedürfen wir einer neuen Beratungskultur, innerhalb derer auch alternative Lebensentwürfe und Wege beruflicher Verwirklichung in neu entstehenden Berufsfeldern jenseits herkömmlicher Vorstellungen von Karrierechancen und Ausbildungswegen reflektierbar werden. Nur so erscheint es möglich, eben nicht mehr der ‚Digitalisierung‘ eine Verantwortlichkeit für gesellschaftliche Umbrüche zuzuschreiben, die doch bei der Gesellschaft selbst verbleibt, sofern diese in der Lage ist, ihre Kräfte zu koordinieren und

nicht einfach einem Machtgeschehen zu überlassen, als dessen verantwortlicher Träger dann eine ‚maschinelle Intelligenz‘ erscheint.

LITERATUR

- Abowd, Gregory D./Dey, Anind K./Brown, Peter J./Davies, Nigel/Smith, Mark/Steggles, Pete (1999): „Towards a Better Understanding of Context and Context-Awareness“, in: Hans-Werner Gellersen (Hg.), *Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing* (= *Lecture Notes in Computer Science* 1707), London: Springer, S. 304-307.
- Alpaydin, Ethem (2008): *Maschinelles Lernen*, München: Oldenbourg.
- Braun-Thürmann, Holger (2002): *Künstliche Interaktion. Wie Technik zur Teilnehmerin sozialer Wirklichkeit wird*, Wiesbaden: Westdeutscher Verlag.
- Chalmers, Matthew/Maccoll, Ian (2003): „Seamful and Seamless Design in Ubiquitous Computing“, in: *Proceedings of Workshop at the Crossroads: The Interaction of HCI and Systems Issue in UbiComp*, http://www.academia.edu/2456375/Seamful_and_Seamless_Design_in_Ubiquitous_Computing vom 23.8.2018.
- EU-Kommission (2017): *Bericht mit Empfehlungen an die Kommission zu zivilrechtlichen Regelungen im Bereich Robotik*, A8-0005/2017 vom 27.1.2017, <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2017-0005+0+DOC+XML+V0//DE> vom 22.8.2018.
- Feiner, Steven K. (2002): „Augmented Reality: A New Way of Seeing“, in: *Scientific American* 286/4, S. 48-55.
- Fink, Robin D. (2009): *Attributionsprozesse in hybriden Systemen. Experimentelle Untersuchung des Zusammenspiels von Mensch und autonomer Technik* (= *Soziologisches Arbeitspapier* 25), Dortmund: Technische Universität Dortmund.

- Günther, Jan-Philipp (2016): *Roboter und rechtliche Verantwortung. Eine Untersuchung der Benutzer- und Herstellerhaftung*, München: Herbert Utz (insbes. S. 237-255).
- Hubig, Christoph (2007): *Die Kunst des Möglichen 2. Ethik der Technik als provisorische Moral*, Bielefeld: transcript.
- Hubig, Christoph (2015): *Die Kunst des Möglichen 3. Macht der Technik*, Bielefeld: transcript.
- Kotsiantis, Sotiris B. (2007): „Supervised Machine Learning: A Review of Classical Techniques“, in: *Informatica* 31/2007, S. 249-268.
- Krummheuer, Antonia (2010): *Interaktion mit virtuellen Agenten? Zur Aneignung eines ungewohnten Artefakts*, Stuttgart: Lucius & Lucius.
- Lake, Brenden M./Ullmann, Tomer D./Tenenbaum, Joshua B./Gershman, Samuel J. (2016): „Building Machines That Learn and Think Like People“, in: *Behavioral and Brain Sciences*, S. 1-101.
- Latour, Bruno (2006a): „Sozialtheorie und die Erforschung computerisierter Arbeitsumgebungen“, in: Andréa Belliger/David J. Krieger (Hg.), *ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie*, Bielefeld: transcript, S. 561-572.
- Latour, Bruno (2006b): „Über technische Vermittlung: Philosophie, Soziologie und Genealogie“, in: Andréa Belliger/David J. Krieger (Hg.), *ANThology. Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie*, Bielefeld: transcript, S. 483-528.
- Luhmann, Niklas (1984): *Soziale Systeme. Grundriss einer allgemeinen Theorie*, Frankfurt a.M.: Suhrkamp.
- Rammert, Werner/Schulz-Schaeffer, Ingo (2002): „Technik und Handeln. Wenn soziales Handeln sich auf menschliches Verhalten und technische Artefakte verteilt“, in: dies. (Hg.), *Können Maschinen handeln? Soziologische Beiträge zum Verhältnis von Mensch und Technik*, Frankfurt a.M. u.a.: Campus, S. 11-64.
- Schilling, Malte/Kopp, Stefan/Wachsmuth, Sven/Wrede, Britta/Ritter, Helge (2016): „Towards a Multidimensional Perspective on Shared

- Autonomy“, in: The 2016 AAAI Fall Symposium Series: Shared Autonomy in Research and Practice Technical Report, S. 338-344.
- Weber, Max (1968): *Gesammelte Aufsätze zur Wissenschaftslehre*, Tübingen: Mohr Siebeck.
- Weyer, Johannes/Fink, Robin D. (2011): „Die Interaktion von Mensch und autonomer Technik in soziologischer Perspektive“, in: *Technikfolgenabschätzung – Theorie und Praxis*, 20/1, S. 39-45.
- Wiegerling, Klaus (2011): *Philosophie intelligenter Welten*, Paderborn: Wilhelm Fink.

Autorinnen und Autoren

Brödner, Peter, Prof. Dr.-Ing., Honorarprofessor für Wirtschaftsinformatik, Universität Siegen; Arbeitsgebiete: Theorie der Informatik; Gestaltung computerunterstützter Arbeit und Organisation.

Gerst, Detlef, Dr., Ressortleiter Zukunft der Arbeit, IG Metall Vorstand; Arbeitsgebiete: Digitalisierung von Arbeit, Produktionssysteme, Arbeitsgestaltung.

Herrmann, Thomas, Prof. Dr., Professor für Informations- und Technikmanagement am Institut für Arbeitswissenschaft der Ruhr-Universität Bochum (RUB); Arbeitsgebiete: Soziotechnisches Design im Bereich Industrie 4.0, Wissens- und Prozessmanagement, Gesundheitsversorgung und Kreativitätsförderung.

Heßler, Martina, Prof. Dr., Professorin für neuere Sozial-, Wirtschafts- und Technikgeschichte an der Helmut-Schmidt-Universität Hamburg; Arbeitsgebiete: Technikgeschichte des 20. Jahrhunderts, Historische Technikanthropologie, Mensch-Maschinen-Verhältnisse seit der Frühen Neuzeit.

Hirsch-Kreinsen, Hartmut, Prof. Dr., Professor i.R. für Wirtschafts- und Industriesoziologie an der Technischen Universität Dortmund; Arbeitsgebiete: Entwicklung von Arbeit, Innovationsprozesse und Tech-

nologieentwicklung sowie insbesondere Digitalisierung der industriellen Produktion.

Hubig, Christoph, Prof. Dr., em., Professor für Praktische Philosophie/Philosophie der wiss.-techn. Kultur, Institut für Philosophie der Technischen Universität Darmstadt; Arbeitsgebiete: Technikphilosophie, Wissenschaftstheorie, anwendungsbezogene Ethik, Sozialphilosophie.

Huchler, Norbert, Dr., Wissenschaftler und Mitglied des Vorstands am Institut für Sozialwissenschaftliche Forschung München e.V. (ISF München); Arbeitsgebiete: Soziale Mechanismen, Koordination und Steuerung von Arbeit, erfahrungsgeleitet-subjektivierendes Arbeits-handeln, Verhältnis Mensch-Technik-Organisation, digitaler Wandel.

Igel, Christoph, Prof. Dr., Wissenschaftlicher Direktor, Educational Technology Lab des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI) Berlin; Arbeitsgebiete: Bildungstechnologien, technologiebasierte Qualifizierung und Arbeit, Digitalisierung, Industrie 4.0 und Lernende Systeme.

Karačić, Anemari, Wissenschaftliche Referentin am Forschungsinstitut für gesellschaftliche Weiterentwicklung (FGW); Arbeitsgebiete: Digitalisierung von Arbeit, Transformation der Arbeitswelt, Disability Studies.

Kravčik, Miloš, Dr., Senior Researcher, Educational Technology Lab des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI) Berlin; Arbeitsgebiete: Lerntechnologien, lebenslange Kompetenzentwicklung; mobiles und spielbasiertes Lernen, arbeitsplatzintegriertes Training.

Kuhn, Thomas, Dr., Hauptabteilungsleiter Eingebettete Systeme, Fraunhofer Institut für Experimentelles Software Engineering (IESE);

Arbeitsgebiete: Virtuelle Entwicklung, Software-Architekturen und Industrie 4.0.

Liggesmeyer, Peter, Prof. Dr.-Ing., Geschäftsführender Direktor, Fraunhofer Institut für Experimentelles Software Engineering (IESE); Arbeitsgebiete: Systems Engineering für Cyber-Physische Systeme mit Schwerpunkt auf Sicherheit, Zuverlässigkeit und Verfügbarkeit.

Nierhoff, Jan, Dr., Wissenschaftlicher Mitarbeiter, Lehrstuhl für Informations- und Technikmanagement der Ruhr-Universität Bochum (RUB); Arbeitsgebiete: Gestaltungskriterien für soziotechnische Systeme, Innovationsprozesse im Kontext digitaler Transformation, Gestaltung kollaborativer Systeme für selbstbestimmte agile Teams, Nutzerzentrierte Softwareentwicklung.

Schulz-Schaeffer, Ingo, Prof. Dr., Leiter des Fachgebiets Technik- und Innovationssoziologie, Institut für Soziologie, Technische Universität Berlin; Arbeitsgebiete: Techniksoziologie, sozialwissenschaftliche Innovationsforschung, Handlungstheorie.

Ullrich, Carsten, Dr.-Ing., stellvertretender Leiter, Educational Technology Lab des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI) Berlin; Arbeitsgebiete: technologiegestütztes Lernen und Arbeiten, Adaptivität und Lernerunterstützung, webbasiertes und mobiles Lernen.

Soziologie



Sighard Neckel, Natalia Besedovsky, Moritz Boddenberg,
Martina Hasenfratz, Sarah Miriam Pritz, Timo Wiegand

Die Gesellschaft der Nachhaltigkeit Umriss eines Forschungsprogramms

Januar 2018, 150 S., kart.

14,99 € (DE), 978-3-8376-4194-3

E-Book kostenlos erhältlich als Open-Access-Publikation

PDF: ISBN 978-3-8394-4194-7

EPUB: ISBN 978-3-7328-4194-3



Sabine Hark, Paula-Irene Villa

Unterscheiden und herrschen

Ein Essay zu den ambivalenten
Verflechtungen von Rassismus,
Sexismus und Feminismus
in der Gegenwart

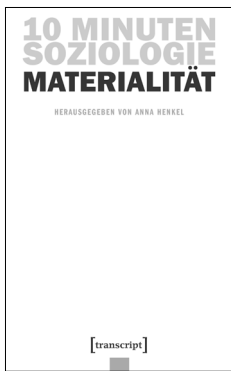
2017, 176 S., kart.

19,99 € (DE), 978-3-8376-3653-6

E-Book

PDF: 17,99 € (DE), ISBN 978-3-8394-3653-0

EPUB: 17,99 € (DE), ISBN 978-3-7328-3653-6



Anna Henkel (Hg.)

10 Minuten Soziologie: Materialität

Juni 2018, 122 S., kart.

15,99 € (DE), 978-3-8376-4073-1

E-Book: 13,99 € (DE), ISBN 978-3-8394-4073-5

**Leseproben, weitere Informationen und Bestellmöglichkeiten
finden Sie unter www.transcript-verlag.de**

Soziologie



Robert Seyfert, Jonathan Roberge (Hg.)

Algorithmuskulturen

Über die rechnerische Konstruktion der Wirklichkeit

2017, 242 S., kart., Abb.

29,99 € (DE), 978-3-8376-3800-4

E-Book kostenlos erhältlich als Open-Access-Publikation

PDF: ISBN 978-3-8394-3800-8

EPUB: ISBN 978-3-7328-3800-4



Andreas Reckwitz

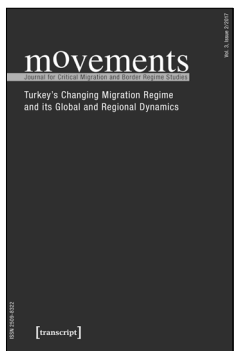
Kreativität und soziale Praxis

Studien zur Sozial- und Gesellschaftstheorie

2016, 314 S., kart.

29,99 € (DE), 978-3-8376-3345-0

E-Book: 26,99 € (DE), ISBN 978-3-8394-3345-4



Ilker Ataç, Gerda Heck, Sabine Hess, Zeynep Kasli,
Philipp Ratfisch, Cavidan Soykan, Bediz Yilmaz (eds.)

movements. Journal for Critical Migration and Border Regime Studies

Vol. 3, Issue 2/2017:

Turkey's Changing Migration Regime
and its Global and Regional Dynamics

2017, 230 p., pb.

24,99 € (DE), 978-3-8376-3719-9

**Leseproben, weitere Informationen und Bestellmöglichkeiten
finden Sie unter www.transcript-verlag.de**

